Canella, K.
08/731499
Seq. 109

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 02:36:38 ; Search time 2069 Seconds
(without alignments)
8548.165 Million cell updates/sec

Title: US-08-731-499-9_COPY_10001_10365
Perfect score: 365
Sequence: 1 TTTGTGGTCCTCCAAGGCTT.....GATGCACTCCACCACGCTTG 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	365	100.0	10365	6	BD085733	BD085733 Genes fro
c	2	365	100.0	121143	9	AF312915	Homo sapi
c	3	365	100.0	128871	9	AL157838	Human DNA
	4	363.4	99.6	105023	2	AC116668	Trypanoso
	5	164	44.9	109710	9	HSBA46E17	Human DNA
c	6	159.2	43.6	165365	2	AC011959	Homo sapi
c	7	159.2	43.6	184851	9	AC100809	Homo sapi
	8	159.2	43.6	321261	9	AF131216	Homo sapi
c	9	153.8	42.1	188141	2	AC023120	Homo sapi
	10	153.4	42.0	192499	2	AC133913	Homo sapi
	11	152.6	41.8	117230	9	HSJ684F13	Human DNA
	12	152.2	41.7	168495	2	AC084299	Homo sapi
c	13	152	41.6	162216	2	AC136144	Pan trogl
	14	151.4	41.5	151408	2	AC010397	Homo sapi
	15	151.4	41.5	155046	2	AC146346	Pan trogl
	16	151.4	41.5	182368	2	AC145822	Pan trogl
c	17	151.4	41.5	184010	2	AC150025	Papio anu
c	18	151.4	41.5	196473	2	AC146345	Pan trogl
	19	151.4	41.5	253038	2	AC008930	Homo sapi
c	20	151	41.4	182509	9	AC112211	AC112211 Homo sapi
c	21	150.2	41.2	157267	9	AC104938	AC104938 Homo sapi
	22	150.2	41.2	170388	2	AC027750	AC027750 Homo sapi
	23	150.2	41.2	203530	9	AC025097	AC025097 Homo sapi
c	24	149.4	40.9	180303	9	AL672045	AL672045 Human DNA
	25	149.4	40.9	201012	2	AC021189	AC021189 Homo sapi
	26	149.2	40.9	167869	9	AC073342	AC073342 Homo sapi
	27	149	40.8	104295	9	AC117502	AC117502 Homo sapi
	28	149	40.8	149202	2	AC022160	AC022160 Homo sapi
c	29	148.8	40.8	102562	2	AC109516	AC109516 Homo sapi
	30	148.8	40.8	119050	9	AC127024	AC127024 Homo sapi
c	31	148.8	40.8	153520	9	AC130324	AC130324 Homo sapi
c	32	148.8	40.8	157502	2	AC145842	AC145842 Papio anu
	33	148.8	40.8	173577	2	AC146322	AC146322 Papio anu
	34	148.8	40.8	180537	2	AC023266	AC023266 Homo sapi
c	35	148.8	40.8	192030	9	AC092406	AC092406 Papio anu
	36	148.6	40.7	160099	2	AC145498	AC145498 Papio anu
c	37	148.4	40.7	187738	9	AL451140	AL451140 Human DNA
	38	148.4	40.7	189456	9	AC008509	AC008509 Homo sapi
	39	148	40.5	189174	9	AC092687	AC092687 Homo sapi
c	40	147.6	40.4	83661	9	AP001439	AP001439 Homo sapi
c	41	147.6	40.4	100000	9	AP000143	AP000143 Homo sapi
	42	147.6	40.4	110000	2	AC143327_2	Continuation (3 of
c	43	147.6	40.4	121597	9	AP000090	AP000090 Homo sapi
c	44	147.6	40.4	168778	9	HSA490O24	AL117338 Human DNA
c	45	147.6	40.4	189256	2	AL391004	AL391004 Homo sapi

ALIGNMENTS

RESULT 1
BD085733
LOCUS BD085733 10365 bp DNA linear PAT 27-AUG-2002
DEFINITION Genes from the 20q13 amplicon and their uses.
ACCESSION BD085733
VERSION BD085733.1 GI:22631343
KEYWORDS JP 2001524802-A/9.
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 10365)
AUTHORS Gray,J.W., Collins,C.C., Hwang,S.I., Godfrey,T., Kowbel,D. and Rommens,J.
TITLE Genes from the 20q13 amplicon and their uses
JOURNAL Patent: JP 2001524802-A 9 04-DEC-2001;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
COMMENT OS Artificial Sequence
PN JP 2001524802-A/9
PD 04-DEC-2001
PF 15-JUL-1997 JP 1998506264
PR 15-JUL-1996 US 08/680395,16-OCT-1996 US 08/731499 PR 17-JAN-1997 US 08/785532
PI JOE W GRAY,COLIN CONRAD COLLINS,SOO IN HWANG,TONY GODFREY, PI DAVID KOWBEL,
PI JOHANNA ROMMENS
PC C12N15/11,C12Q1/68,A61K48/00
CC Description of Artificial Sequence:Genomic Sequence encoding
CC ZABC1
FH Key Location/Qualifiers
FT source 1. .10365
FT /organism='Artificial Sequence'.
FEATURES Location/Qualifiers
source 1. .10365
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN

Query Match 100.0%; Score 365; DB 6; Length 10365;
Best Local Similarity 100.0%; Pred. No. 3.4e-90;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10001 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 10060

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10061 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 10120

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 10180

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 10240

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 10300

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 10360

Qy        361 GCTTG 365
              |||||
Db      10361 GCTTG 10365
```

RESULT 2
AF312915/c
LOCUS AF312915 121143 bp DNA linear PRI 13-JUN-2001
DEFINITION Homo sapiens chromosome 20 clones 97 and 127, complete sequence.
ACCESSION AF312915
VERSION AF312915.1 GI:11094030
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 121143)
AUTHORS Collins,C., Volik,S., Kowbel,D., Ginzinger,D., Ylstra,B., Cloutier,T., Hawkins,T., Predki,P., Martin,C., Wernick,M., Kuo,W.L., Alberts,A. and Gray,J.W.
TITLE Comprehensive genome sequence analysis of a breast cancer amplicon
JOURNAL Genome Res. 11 (6), 1034-1042 (2001)
MEDLINE 21275464
PUBMED 11381030
REFERENCE 2 (bases 1 to 121143)
AUTHORS Volik,S., Collins,C., Gray,J., Wernick,M., Kowbel,D., Stultz,K. and Martin,C.
TITLE Direct Submission
JOURNAL Submitted (10-OCT-2000) Cancer Genetics, UCSF Cancer Center, 2340 Sutter St., Rm. S151, San Francisco, CA 94706, USA

FEATURES Location/Qualifiers
source 1. .121143
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.2"
/clone="97; 127; LBNL h117 and h79"
/note="isolated in Collins lab and sequenced by the Genome Center at LBL"
misc_feature complement(1. .66093)
/note="overlaps Homo sapiens chromosome 20 clone RP5-823G15 deposited in GenBank Accession Number AL354993.19"
repeat_region 194. .468
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region complement(513. .653)
/rpt_family="MER5"
/rpt_type=dispersed
repeat_region 1053. .1153
/rpt_family="MER20"
/rpt_type=dispersed
repeat_region 1254. .1330
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region complement(1356. .1676)
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 1681. .1866
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature 2485. .5421
/note="similar to Homo sapiens chromosome 5 clone CTC-269F1 deposited in GenBank Accession Number AC011341.2"
repeat_region 2594. .2879
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature complement(2896. .4437)
/note="similar to flow-sorted chromosome 20 HindIII fragment, SC20pF17C6 deposited in GenBank Accession Number Z94682.1"
repeat_region 3106. .3412
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 3889. .4206
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 4449. .4764
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region complement(5671. .5987)
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 6723. .7004
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region complement(7486. .7625)
/rpt_family="L1"
/rpt_type=dispersed
repeat_region 7716. .7965
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature complement(8590. .118606)
/note="similar to Mus musculus clone CT7-510C4 deposited in GenBank Accession Number AC023610.3"
misc_feature complement(9852. .10109)
/note="similar to STS WI-31010 deposited in GenBank Accession Number G24374.1"
mRNA <10244. .>10904
/product="hypothetical"
/note="similar to Homo sapiens hypothetical protein OVC10-2 deposited in GenBank Accession Numbers AF230201.1, AP000365 and in NM_018692.1"
/evidence=not_experimental
repeat_region 15301. .15588
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature 16873. .17073
/note="similar to Homo sapiens STS WI-16697 deposited in GenBank Accession Number G21490.1"
repeat_region 17577. .17857
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region complement(19034. .19333)
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 20136. .20719
/rpt_family="Alu"
/rpt_type=dispersed
repeat_region 20776. .21071

```
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(22344. .22638)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   22726. .22784
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(22984. .23268)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(23349. .23646)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(24008. .24189)
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   complement(24502. .24795)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   24883. .24941
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(25140. .25425)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(25507. .25805)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(26170. .26352)
                     /rpt_family="MIR"
                     /rpt_type=dispersed
     repeat_region   26841. .27129
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   27270. .33290
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   28512. .28730
                     /rpt_family="MER25"
                     /rpt_type=dispersed
     repeat_region   34498. .35009
                     /rpt_family="MLT2B2"
                     /rpt_type=dispersed
     repeat_region   35035. .35332
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   35343. .35522
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     misc_feature    35593. .121143
                     /note="overlaps Homo sapiens clone RP4-724E16 on
                     chromosome 20 deposited in GenBank Accession Number
                     AL157838.24"
     repeat_region   36137. .36250
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   36366. .36690
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   36949. .37020
                     /rpt_family="MER5"
                     /rpt_type=dispersed
     repeat_region   complement(37692. .37766)
                     /rpt_family="MER5"
                     /rpt_type=dispersed
     repeat_region   complement(38006. .38276)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   38863. .38984
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(39791. .40072)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(40100. .40395)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   40642. .41118
                     /rpt_family="MER1"
                     /rpt_type=dispersed
     repeat_region   complement(41228. .41348)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   41649. .41938

  Query Match             100.0%;  Score 365;  DB 9;  Length 121143;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-90;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97755 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 97696

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97695 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 97636

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97635 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 97576

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97575 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 97516

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97515 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 97456

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97455 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 97396

Qy        361 GCTTG 365
              |||||
Db      97395 GCTTG 97391


RESULT 3
AL157838/c
LOCUS       AL157838               128871 bp    DNA     linear   PRI 23-FEB-2001
DEFINITION  Human DNA sequence from clone RP4-724E16 on chromosome
            20q13.12-13.32 Contains the ZNF217 gene for zinc finger protein
            217, a novel gene, a putative novel gene, ESTs, GSSs, STSs and two
            CpG islands, complete sequence.
ACCESSION   AL157838
VERSION     AL157838.24  GI:9588158
KEYWORDS    HTG; CpG island; zinc finger; ZNF217.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128871)
  AUTHORS   Wilson,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 31, 2000 this sequence version replaced gi:9408255.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
```

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr20
This sequence is the entire insert of clone RP4-724E16 The true right end of clone RP5-823G15 is at 19684 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP4-724E16 is from the library RPCI-4 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.

FEATURES Location/Qualifiers

source 1. .128871
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="q13.12-13.32"
/clone="RP4-724E16"
/clone_lib="RPCI-4"

repeat_region 1. .54
/note="MLT2A repeat: matches 362. .415 of consensus"

repeat_region 55. .368
/note="AluSx repeat: matches 1. .312 of consensus"

repeat_region 369. .559
/note="AluSq repeat: matches 126. .312 of consensus"

repeat_region 1137. .1270
/note="FLAM_C repeat: matches 1. .133 of consensus"

misc_feature complement(1396. .1840)
/note="match: GSS: Em:AQ261624"

repeat_region 1415. .1709
/note="AluJo repeat: matches 14. .288 of consensus"

repeat_region 1744. .1902
/note="MIR repeat: matches 14. .175 of consensus"

misc_feature 1871. .2326
/note="match: GSS: Em:AQ673688"

repeat_region 1948. .2041
/note="MER5A repeat: matches 6. .101 of consensus"

repeat_region 2622. .2787
/note="MER5A repeat: matches 4. .189 of consensus"

repeat_region 2806. .2870
/note="MER5A repeat: matches 10. .75 of consensus"

repeat_region 2872. .3026
/note="L1MA4 repeat: matches 6017. .6162 of consensus"

repeat_region 3027. .3307
/note="AluJo repeat: matches 1. .290 of consensus"

repeat_region 3308. .3436
/note="L1MA4 repeat: matches 6162. .6293 of consensus"

repeat_region 3455. .3551
/note="MER5A repeat: matches 61. .183 of consensus"

repeat_region 3564. .3697
/note="MER5B repeat: matches 14. .146 of consensus"

repeat_region 3871. .4004
/note="AluSq/x repeat: matches 1. .134 of consensus"

repeat_region 4249. .4274
/note="13 copies 2 mer gt 96% conserved"

repeat_region 4810. .5101
/note="AluSx repeat: matches 1. .297 of consensus"

repeat_region 5110. .5413
/note="AluSc repeat: matches 1. .305 of consensus"

repeat_region 5492. .5648
/note="L2 repeat: matches 2589. .2745 of consensus"

repeat_region 5659. .6135
/note="MER1A repeat: matches 1. .488 of consensus"

repeat_region 6164. .6225
/note="31 copies 2 mer aa 71% conserved"

repeat_region 6245. .6379
/note="AluSg/x repeat: matches 162. .296 of consensus"

repeat_region 6380. .6537
/note="79 copies 2 mer aa 60% conserved"

repeat_region 6658. .6967
/note="AluSg repeat: matches 1. .309 of consensus"

repeat_region 7320. .7542
/note="MER58B repeat: matches 117. .341 of consensus"

repeat_region 7545. .7838
/note="AluSx repeat: matches 1. .293 of consensus"

misc_feature 8465. .9068
/note="match: GSS: Em:AQ386102"

repeat_region 8523. .8814
/note="AluJb repeat: matches 1. .294 of consensus"

repeat_region 9262. .9561
/note="AluSq repeat: matches 1. .301 of consensus"

repeat_region 9562. .9599
/note="19 copies 2 mer tt 81% conserved"

repeat_region 9606. .9733
/note="MER33 repeat: matches 17. .148 of consensus"

repeat_region 9761. .9915
/note="MER5A repeat: matches 4. .183 of consensus"

repeat_region 9980. .10108
/note="L2 repeat: matches 2622. .2747 of consensus"

repeat_region 10128. .10437
/note="AluSx repeat: matches 2. .312 of consensus"

repeat_region 10654. .10724
/note="MIR repeat: matches 63. .137 of consensus"

repeat_region 10781. .11083
/note="AluSx repeat: matches 1. .306 of consensus"

misc_feature complement(10897. .11319)
/note="match: GSS: Em:AQ499673"

misc_feature complement(10991. .11319)
/note="match: GSS: Em:AQ283579"

repeat_region 11462. .11759
/note="AluSq repeat: matches 1. .299 of consensus"

repeat_region 11878. .12182
/note="AluSx repeat: matches 1. .300 of consensus"

repeat_region 12183. .12587
/note="L1PA8 repeat: matches 5757. .6163 of consensus"

repeat_region 13333. .13637
/note="MLT1G repeat: matches 3. .338 of consensus"

repeat_region 14201. .14507
/note="AluY repeat: matches 1. .308 of consensus"

repeat_region 14540. .14694
/note="L2 repeat: matches 2550. .2709 of consensus"

repeat_region 14764. .14868
/note="MIR repeat: matches 103. .208 of consensus"

repeat_region 14891. .15152
/note="AluJo repeat: matches 12. .299 of consensus"

repeat_region 15162. .15293
/note="L2 repeat: matches 1697. .1824 of consensus"

repeat_region 15317. .15544
/note="L1MB3 repeat: matches 5943. .6173 of consensus"

repeat_region 16306. .16603
/note="AluSc repeat: matches 1. .306 of consensus"

repeat_region 16610. .16901
/note="AluSx repeat: matches 1. .292 of consensus"

repeat_region 16989. .17035
/note="MLT1C repeat: matches 420. .463 of consensus"

repeat_region 17036. .17338
/note="AluJo repeat: matches 1. .301 of consensus"

repeat_region 17339. .17730
/note="MLT1C repeat: matches 1. .420 of consensus"

repeat_region 17958. .18269
/note="AluSx repeat: matches 1. .311 of consensus"

repeat_region 18288. .18315
/note="MER91 repeat: matches 32. .59 of consensus"

repeat_region 18383. .18507

```
                     /note="MIR repeat: matches 23. .148 of consensus"
     repeat_region   18544. .18650
                     /note="MIR repeat: matches 137. .261 of consensus"
     repeat_region   18799. .18820
                     /note="11 copies 2 mer aa 100% conserved"
     repeat_region   18835. .19184
                     /note="THE1A repeat: matches 1. .354 of consensus"
     misc_feature    19231. .19522
                     /note="match: STS: Em:HSA153YF5"
     repeat_region   19297. .19346
                     /note="25 copies 2 mer ca 100% conserved"
     repeat_region   19426. .19564
                     /note="MIR repeat: matches 6. .147 of consensus"
     repeat_region   19624. .19916
                     /note="AluSx repeat: matches 1. .295 of consensus"
     repeat_region   20005. .20204
                     /note="MER20 repeat: matches 1. .218 of consensus"
     repeat_region   20215. .20317
                     /note="MER5A repeat: matches 4. .108 of consensus"
     repeat_region   20410. .20541
                     /note="MER5B repeat: matches 1. .120 of consensus"
     repeat_region   20544. .20687
                     /note="AluSg/x repeat: matches 154. .297 of consensus"
     repeat_region   20689. .20814
                     /note="FLAM_A repeat: matches 14. .136 of consensus"
     repeat_region   20816. .21218
                     /note="Charlie4a repeat: matches 17. .427 of consensus"
     repeat_region   21337. .21634
                     /note="AluY repeat: matches 1. .297 of consensus"

  Query Match             100.0%;  Score 365;  DB 9;  Length 128871;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-90;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64003 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 63944

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63943 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 63884

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63883 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 63824

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63823 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 63764

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63763 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 63704

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63703 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 63644

Qy        361 GCTTG 365
              |||||
Db      63643 GCTTG 63639


RESULT 4
AC116668
LOCUS       AC116668              105023 bp    DNA     linear   HTG 01-MAY-2002
DEFINITION  Trypanosoma brucei chromosome V clone RPCI93-45E22, *** SEQUENCING
            IN PROGRESS ***, 3 unordered pieces.
ACCESSION   AC116668
VERSION     AC116668.5  GI:20376999
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Trypanosoma brucei
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 105023)
  AUTHORS   El-Sayed,N.M., Ghedin,E., Song,J., Larkin,C., Wanless,D., Jones,K.,
            Peterson,J., Hou,L., Zhao,H., Mason,T., Militscher,J., Pai,G., Van
            Aken,S., Utterback,T., Khalak,H.G., Gerard,C., Leech,V., Ullu,E.,
            Melville,S., White,O., Adams,M.D., Donelson,J.E. and Fraser,C.M.
  TITLE     Trypanosoma brucei GUTat10.1 RPCI93-45E22 BAC genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105023)
  AUTHORS   El-Sayed,N.M., Khalak,H. and Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 105023)
  AUTHORS   El-Sayed,N.M., Khalak,H. and Adams,M.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On May 1, 2002 this sequence version replaced gi:20340472.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    85781: contig of 85781 bp in length
            *   85782    85806: gap of unknown length
            *   85807   102956: contig of 17150 bp in length
            *  102957   102981: gap of unknown length
            *  102982   105023: contig of 2042 bp in length.
FEATURES             Location/Qualifiers
     source          1. .105023
                     /organism="Trypanosoma brucei"
                     /mol_type="genomic DNA"
                     /isolate="GUTat10.1"
                     /db_xref="taxon:5691"
                     /chromosome="V"
                     /clone="RPCI93-45E22"
ORIGIN

  Query Match             99.6%;  Score 363.4;  DB 2;  Length 105023;
  Best Local Similarity   99.7%;  Pred. No. 9.4e-90;
  Matches  364;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14284 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 14343

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14344 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 14403

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14404 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 14463

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      14464 GATCCTTCTGCCTCAGCTTCCCAGGTAGCGGAGACTATATGTGCTGCTACCATGCACAGC 14523

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14524 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 14583

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      14584 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 14643

Qy        361 GCTTG 365
              |||||
```

```
Db      14644 GCTTG 14648


RESULT 5
HSBA46E17
LOCUS       HSBA46E17             109710 bp    DNA     linear   PRI 02-DEC-2000
DEFINITION  Human DNA sequence from clone RP11-46E17 on chromosome 22, complete
            sequence.
ACCESSION   AL050402
VERSION     AL050402.16  GI:5832404
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109710)
  AUTHORS   Collier,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 6, 1999 this sequence version replaced gi:5791503.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            has been finished according to sequence map criteria as follows.
            An attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            RP11-46E17 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-46E17 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP1-205F14p is at 109611 in this
            sequence. The true right end of clone RP1-231P7p is at 100 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .109710
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="RP11-46E17"
                     /clone_lib="RPCI-11.1"
     repeat_region   103. .416
                     /note="AluSg1 repeat: matches 1. .306 of consensus"
     repeat_region   4948. .5067
                     /note="MER45 repeat: matches 1. .120 of consensus"
     repeat_region   6308. .6764
                     /note="L1MB5 repeat: matches 5577. .6034 of consensus"
     misc_feature    9512. .9946
                     /note="match: GSS: Em:AQ900161"
     misc_feature    9706. .10111
                     /note="match: STS: Em:HSB294ZC1"
     repeat_region   9862. .9897
                     /note="18 copies 2 mer tc 94% conserved"
     repeat_region   11101. .11201
                     /note="MER45 repeat: matches 83. .178 of consensus"
     repeat_region   11498. .11578
                     /note="MER45 repeat: matches 1. .83 of consensus"
     misc_feature    13423. .13654
                     /note="match: GSS: Em:AQ490130"
     misc_feature    complement(14203. .14663)
                     /note="match: GSS: Em:AQ613979"
     misc_feature    14712. .15061
                     /note="match: STS: Em:G50283
                     match: GSS: Em:AQ044902"
     misc_feature    complement(17350. .17747)
                     /note="match: GSS: Em:AQ534637"
     misc_feature    complement(17374. .17760)
                     /note="match: GSS: Em:AQ179064"
     misc_feature    17772. .18403
                     /note="match: GSS: Em:AQ508370"
     misc_feature    17795. .18238
                     /note="match: GSS: Em:AQ268578"
     repeat_region   19884. .19939
                     /note="28 copies 2 mer ag 96% conserved"
     repeat_region   24162. .24201
                     /note="20 copies 2 mer gt 92% conserved"
     misc_feature    30439. .30918
                     /note="match: GSS: Em:AQ709314"
     repeat_region   30586. .30896
                     /note="AluYa5 repeat: matches 1. .311 of consensus"
     repeat_region   31362. .31427
                     /note="33 copies 2 mer gg 66% conserved"
     misc_feature    complement(33542. .33960)
                     /note="match: GSS: Em:AQ416617"
     repeat_region   34761. .34816
                     /note="L1PA5 repeat: matches 5819. .5875 of consensus"
     repeat_region   34817. .35099
                     /note="L1PA5 repeat: matches 5861. .6143 of consensus"
     misc_feature    37097. .37433
                     /note="match: GSS: Em:AQ136413"
     repeat_region   41848. .42021
                     /note="FRAM repeat: matches 1. .165 of consensus"
     misc_feature    complement(42890. .43566)
                     /note="match: GSS: Em:AQ895335"
     misc_feature    complement(43186. .43566)
                     /note="match: GSS: Em:AQ166907"
     repeat_region   43363. .43781
                     /note="L1MD2 repeat: matches 5929. .7739 of consensus"
     misc_feature    43575. .44111
                     /note="match: GSS: Em:AQ376177"
     repeat_region   50205. .50541
                     /note="MER58B repeat: matches 1. .341 of consensus"
     repeat_region   50755. .50870
                     /note="L1ME3A repeat: matches 6050. .6163 of consensus"
     repeat_region   54192. .54655
                     /note="232 copies 2 mer aa 56% conserved"
     misc_feature    58652. .59429
                     /note="match: GSS: Em:AQ739794"
     misc_feature    61986. .62254
                     /note="match: GSS: Em:AL219668"
     misc_feature    complement(62014. .62165)
                     /note="match: GSS: Em:AL253608"
     misc_feature    62025. .62196
                     /note="match: GSS: Em:AQ929489"
     misc_feature    complement(62073. .62436)
                     /note="match: GSS: Em:AL003813"
     misc_feature    62185. .62463
                     /note="match: GSS: Em:AZ078596"
     misc_feature    complement(62204. .62509)
                     /note="match: GSS: Em:AL001671"
     misc_feature    complement(62291. .62446)
                     /note="match: STS: Em:G08353"
     misc_feature    complement(62314. .62461)
                     /note="match: GSS: Em:AQ744114"
     misc_feature    complement(join(62322. .62497,79689. .79824))
                     /note="match: STS: Em:L30312"
     misc_feature    join(62324. .62452,76137. .76341)
```

```
                     /note="match: STS: Em:L18062"
     misc_feature    complement(62332. .62473)
                     /note="match: STS: Em:G08433"
     misc_feature    62334. .62586
                     /note="match: GSS: Em:AQ972678"
     misc_feature    62334. .62463
                     /note="match: GSS: Em:AL130784"
     misc_feature    62335. .62469
                     /note="match: GSS: Em:AL201275"
     misc_feature    complement(62335. .62462)
                     /note="match: STS: Em:G15757"
     unsure          62335
                     /note="Tandem repeat"
     misc_feature    62336. .62540
                     /note="match: GSS: Em:AL013940"
     misc_feature    62336. .62531
                     /note="match: GSS: Em:AL013938"
     misc_feature    complement(62336. .62530)
                     /note="match: GSS: Em:AL164076"
     misc_feature    62336. .62509
                     /note="match: GSS: Em:AL013918"
     misc_feature    complement(62336. .62507)
                     /note="match: GSS: Em:AL002631"
     misc_feature    complement(62344. .62470)
                     /note="match: GSS: Em:AL273514"
     misc_feature    complement(62373. .62507)
                     /note="match: STS: Em:L30408"
     misc_feature    complement(62373. .62494)
                     /note="match: STS: Em:L30526"
     misc_feature    62379. .62515
                     /note="match: STS: Em:G08890"
     misc_feature    join(62385. .62479,80291. .80469)
                     /note="match: STS: Em:AU029014"
     misc_feature    62389. .62503
                     /note="match: GSS: Em:AL268743"
     misc_feature    complement(62390. .62608)
                     /note="match: GSS: Em:AL003817"
     misc_feature    62396. .62538
                     /note="match: GSS: Em:AL185783"
     misc_feature    complement(62403. .62489)
                     /note="match: GSS: Em:AL129551 Em:AL129826"
     misc_feature    complement(62408. .62665)
                     /note="match: STS: Em:G05108"
     misc_feature    complement(62409. .62578)
                     /note="match: GSS: Em:Z88547"
     misc_feature    62422. .62512
                     /note="match: STS: Em:G07969"
     misc_feature    complement(join(62424. .62712,80207. .80531))
                     /note="match: GSS: Em:AL193675"
     misc_feature    complement(62449. .62606)
                     /note="match: STS: Em:G08976"
     misc_feature    62496. .62545
                     /note="Tandem repeat. weak data"
     repeat_region   64167. .64280
                     /note="MER45 repeat: matches 10. .123 of consensus"
     repeat_region   66608. .66733
                     /note="63 copies 2 mer tg 61% conserved"
     repeat_region   67482. .67573
                     /note="46 copies 2 mer ta 63% conserved"

  Query Match             44.9%;  Score 164;  DB 9;  Length 109710;
  Best Local Similarity   67.8%;  Pred. No. 1.3e-34;
  Matches  251;  Conservative    0;  Mismatches  105;  Indels   14;  Gaps    1;

Qy         10 CTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTG 69
              |||| ||    |  |  |||     | ||  || |   | || | | | |   | ||||
Db       1095 CTCCCAGCTCCAAATGGCTCCAAATGGTTTTCTTTCTTCTCTCTCTCTCTCTCTTTTTTT 1154

Qy         70 ATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAG 129
                 |||  |||  | ||  ||| ||||| |||| |||||||||| || |||| ||||  |||||
Db       1155 TTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGGGTCTTGTTCTGTCACCTAGGCTGGAG 1214

Qy        130 TGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCT 189
```

```
              || || ||||  ||| || || |||| | ||||||| ||||||| ||||| |||| | ||
Db       1215 TGTAGTGGTGCCATCGTAACTCACTGCAATCTTGAACTCCTGAGCTCAAGTGATCTTCCT 1274

Qy        190 GCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCTGATTTTTA 249
              ||| |||| ||| | ||||||| ||||| | |||| || |||||||   ||| |||||||
Db       1275 GCCCCAGCCTCCTAAGTAGCTGGGACTACAGGTGCCGCCACCATGCCTGGCTAATTTTTA 1334

Qy        250 AATTTTTTTTGTAGAGATGGA-------------GTTGCCCAGGCTGGTCTTGAACTCC 295
              | |||||||| || | |   |               ||||| |||||||||||||||||||
Db       1335 ATTTTTTTTTTTAAATAGAAATGGGGTCCTGCTATGTTGCTCAGGCTGGTCTTGAACTCC 1394

Qy        296 TGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCC 355
              |||||| |  ||||||| ||    || ||||||||     |  || || || || |  ||
Db       1395 TGGCCTCAAATGATCCTTCTTGTCTGGCCTCCCAAAGTGTTGGGATTATAGGTGTAAGCC 1454

Qy        356 ACCACGCTTG 365
              |||| ||| |
Db       1455 ACCATGCTCG 1464
```

```
RESULT 6
AC011959/c
LOCUS       AC011959              165365 bp    DNA     linear   HTG 20-APR-2000
DEFINITION  Homo sapiens clone RP11-110L10, WORKING DRAFT SEQUENCE, 16
            unordered pieces.
ACCESSION   AC011959
VERSION     AC011959.2  GI:7622344
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165365)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-110L10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165365)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 20, 2000 this sequence version replaced gi:6065411.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2302
                Center clone name: 110_L_10
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 113154 bases at least Q40
                Consensus quality: 135711 bases at least Q30
                Consensus quality: 151557 bases at least Q20
```

```
            Insert size: 185000; agarose-fp
            Insert size: 163865; sum-of-contigs
            Quality coverage: 2.8 in Q20 bases; agarose-fp
            Quality coverage: 3.2 in Q20 bases; sum-of-contigs
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 16 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     1204: contig of 1204 bp in length
          *    1205     1304: gap of 100 bp
          *    1305     3250: contig of 1946 bp in length
          *    3251     3350: gap of 100 bp
          *    3351     6924: contig of 3574 bp in length
          *    6925     7024: gap of 100 bp
          *    7025     9765: contig of 2741 bp in length
          *    9766     9865: gap of 100 bp
          *    9866    15163: contig of 5298 bp in length
          *   15164    15263: gap of 100 bp
          *   15264    18014: contig of 2751 bp in length
          *   18015    18114: gap of 100 bp
          *   18115    24325: contig of 6211 bp in length
          *   24326    24425: gap of 100 bp
          *   24426    31010: contig of 6585 bp in length
          *   31011    31110: gap of 100 bp
          *   31111    39129: contig of 8019 bp in length
          *   39130    39229: gap of 100 bp
          *   39230    46121: contig of 6892 bp in length
          *   46122    46221: gap of 100 bp
          *   46222    56362: contig of 10141 bp in length
          *   56363    56462: gap of 100 bp
          *   56463    67851: contig of 11389 bp in length
          *   67852    67951: gap of 100 bp
          *   67952    79662: contig of 11711 bp in length
          *   79663    79762: gap of 100 bp
          *   79763    92911: contig of 13149 bp in length
          *   92912    93011: gap of 100 bp
          *   93012   122796: contig of 29785 bp in length
          *  122797   122896: gap of 100 bp
          *  122897   165365: contig of 42469 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165365
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-110L10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1204
                     /note="assembly_fragment"
     misc_feature    1305. .3250
                     /note="assembly_fragment"
     misc_feature    3351. .6924
                     /note="assembly_fragment"
     misc_feature    7025. .9765
                     /note="assembly_fragment"
     misc_feature    9866. .15163
                     /note="assembly_fragment"
     misc_feature    15264. .18014
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    18115. .24325
                     /note="assembly_fragment"
     misc_feature    24426. .31010
                     /note="assembly_fragment"
     misc_feature    31111. .39129
                     /note="assembly_fragment"
     misc_feature    39230. .46121
                     /note="assembly_fragment"
     misc_feature    46222. .56362
                     /note="assembly_fragment"
     misc_feature    56463. .67851
                     /note="assembly_fragment"
     misc_feature    67952. .79662
                     /note="assembly_fragment"
     misc_feature    79763. .92911
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    93012. .122796
                     /note="assembly_fragment"
     misc_feature    122897. .165365
                     /note="assembly_fragment"
ORIGIN

  Query Match             43.6%;  Score 159.2;  DB 2;  Length 165365;
  Best Local Similarity   72.4%;  Pred. No. 2.8e-33;
  Matches  220;  Conservative    0;  Mismatches   83;  Indels    1;  Gaps    1;

Qy         58 TTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCA 117
              | | || || |  ||  |||  | ||  | |  ||||| ||| ||||||||||||| ||||
Db      64164 TCTTTTTTTGTTTTTGTTTTCTTGTTGTTGTTGTTTTAAAGACAGGGTCTTGCTCTGTCA 64105

Qy        118 TCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCA 177
              || ||  |||| ||||| ||   |||||||||| |||| |  | || |||||||| |||
Db      64104 TCCAGGCTGGAATGCAGTGGCACAATCATAGCTCACTGCAACCATGGATTCCTGAGCTCA 64045

Qy        178 AGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGC-TGCTACCATGCA 236
              ||  ||||| |||||||||| |||   ||||||| ||||| |  |||   | ||||  |
Db      64044 AGCAATCCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTAGAGCTGCACACCACCACACC 63985

Qy        237 CAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCT 296
              |||| |||||||||  |||||| |||||||||  ||||||| ||||| ||||||||||||
Db      63984 TAGCTAATTTTTAAAACTTTTTTATAGAGATGGGGTTGCCCCGGCTGATCTTGAACTCCT 63925

Qy        297 GGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCA 356
              || || | ||| | | ||||| |   ||||||||     ||  || |||||       |||
Db      63924 GGGCTCAAGTGTTTCCCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAACCA 63865

Qy        357 CCAC 360
              ||||
Db      63864 CCAC 63861


RESULT 7
AC100809/c
LOCUS       AC100809              184851 bp    DNA     linear   PRI 12-DEC-2002
DEFINITION  Homo sapiens chromosome 8, clone CTC-493P15, complete sequence.
ACCESSION   AC100809
VERSION     AC100809.9  GI:26553405
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184851)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone CTC-493P15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184851)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
```

```
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 184851)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 184851)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 12, 2002 this sequence version replaced gi:24757047.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21480
                Center clone name: 493_P_15
            --------------.
FEATURES            Location/Qualifiers
     source         1. .184851
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="8"
                    /map="8"
                    /clone="CTC-493P15"
                    /clone_lib="CITC Human BAC"
     repeat_region  1. .230
                    /rpt_family="MER51A-int"
     repeat_region  232. .596
                    /rpt_family="MER51A"
     repeat_region  597. .886
                    /rpt_family="AluSg"
     repeat_region  887. .1091
                    /rpt_family="MER51A"
     repeat_region  complement(1266. .1328)
                    /rpt_family="L2"
     repeat_region  2361. .2598
                    /rpt_family="L2"
     repeat_region  2638. .2719
                    /rpt_family="(TTTC)n"
     repeat_region  2915. .3051
                    /rpt_family="MER5A"
     repeat_region  3075. .3199
                    /rpt_family="L2"
     repeat_region  complement(3246. .3345)
                    /rpt_family="MIR"
     repeat_region  3352. .3682
                    /rpt_family="MLT1B"
     unsure         3660. .3773
                    /note="single clone coverage"
     unsure         3683. .3688
                    /note="<30 qual single clone coverage"
     repeat_region  complement(3705. .3823)
                    /rpt_family="MIR"
     unsure         3747. .3753
                    /note="<30 qual single clone coverage"
     repeat_region  3859. .4234
                    /rpt_family="THE1C"
     repeat_region  complement(4255. .4484)
                    /rpt_family="MER66B"
     repeat_region  complement(4485. .4813)
                    /rpt_family="AluYb8"
     unsure         complement(4646. .4657)
                    /note="<30 qual SNGL region."
     unsure         complement(4676. .4689)
                    /note="<30 qual SNGL region"
     unsure         complement(4705. .4756)
                    /note="<30 qual SNGL region."
     unsure         complement(4764. .4769)
                    /note="<30 qual SNGL region."
     unsure         complement(4780. .4784)
                    /note="<30 qual SNGL region."
     repeat_region  complement(4814. .5074)
                    /rpt_family="MER66B"
     repeat_region  5123. .5236
                    /rpt_family="L2"
     repeat_region  complement(5358. .5410)
                    /rpt_family="MIR"
     unsure         complement(5879. .5890)
                    /note="single clone coverage"
     repeat_region  complement(6250. .6393)
                    /rpt_family="MLT1G"
     repeat_region  complement(6811. .6982)
                    /rpt_family="AluJb"
     repeat_region  6983. .7039
                    /rpt_family="(TTTTA)n"
     repeat_region  complement(7040. .7158)
                    /rpt_family="AluJb"
     repeat_region  7161. .7216
                    /rpt_family="AT_rich"
     repeat_region  7357. .7424
                    /rpt_family="L1MC4a"
     repeat_region  7958. .8001
```

```
                     /rpt_family="(TATG)n"
     repeat_region   complement(8007. .8407)
                     /rpt_family="MSTA"
     unsure          8015. .8231
                     /note="PCR product only region."
     repeat_region   8408. .8428
                     /rpt_family="(CA)n"
     repeat_region   8436. .8539
                     /rpt_family="L1MC4"
     repeat_region   complement(9212. .9325)
                     /rpt_family="MIR3"
     repeat_region   9438. .9907
                     /rpt_family="MER31A"
     repeat_region   10161. .10205
                     /rpt_family="(TG)n"
     repeat_region   complement(10313. .10590)
                     /rpt_family="MER67B"
     repeat_region   complement(10591. .10894)
                     /rpt_family="AluSq"
     unsure          10639. .10645
                     /note="<30 qual SNGL region."
     repeat_region   complement(10895. .11201)
                     /rpt_family="MER67B"
     unsure          11055. .11061
                     /note="<30 qual SNGL region."
     repeat_region   11353. .11472
                     /rpt_family="MER5A"
     repeat_region   11473. .11794

  Query Match             43.6%;  Score 159.2;  DB 9;  Length 184851;
  Best Local Similarity   72.4%;  Pred. No. 2.8e-33;
  Matches  220;  Conservative    0;  Mismatches   83;  Indels    1;  Gaps    1;

Qy         58 TTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCA 117
              | | || || |  ||  |||  | ||  | |  ||||| ||| ||||||||||| ||||
Db      14898 TCTTTTTTTGTTTTTGTTTTCTTGTTGTTGTTTTAAAGACAGGGTCTTGCTCTGTCA 14839

Qy        118 TCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCA 177
              || ||  |||| ||||| ||   |||||||||| |||| |  | || ||||||||| |||
Db      14838 TCCAGGCTGGAATGCAGTGGCACAATCATAGCTCACTGCAACCATGGATTCCTGAGCTCA 14779

Qy        178 AGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGC-TGCTACCATGCA 236
              ||  ||||| |||||||||| |||   ||||||| ||||| |  |||   | ||||  |
Db      14778 AGCAATCCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTAGAGCTGCACACCACCACACC 14719

Qy        237 CAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCT 296
               |||| |||||||||  |||||| |||||||| |||||||| ||||| ||||||||||
Db      14718 TAGCTAATTTTTAAAACTTTTTTATAGAGATGGGGTTGCCCCGGCTGATCTTGAACTCCT 14659

Qy        297 GGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCA 356
              || || | ||| | | ||||| |   ||||||||    ||  || |||||       |||
Db      14658 GGGCTCAAGTGTTTCCCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAACCA 14599

Qy        357 CCAC 360
              ||||
Db      14598 CCAC 14595


RESULT 8
AF131216
LOCUS       AF131216               321261 bp    DNA     linear   PRI 16-FEB-2004
DEFINITION  Homo sapiens chromosome 8 map 8p23-p22 clone CTB-164D9, CTB-169o5,
            CTB-65D4, CTC-271o23, CTC-367J24, complete sequence; and clones
            CTC-493P15, CTB-91N18, partial sequence.
ACCESSION   AF131216
VERSION     AF131216.2  GI:42557499
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 321261)
  AUTHORS   Appel,S., Reichwald,K., Hennies,H.-C., Zimmermann,W., Reis,A. and
            Rosenthal,A.
  TITLE     Chromosome 8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 321261)
  AUTHORS   Reichwald,K. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 321261)
  AUTHORS   Lagemann,D. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     On Feb 16, 2004 this sequence version replaced gi:8272458.
            -------------- Genome Center
                 Center: Insitute of Molecular Biotechnology
                 Center code: IMB
                 Web site: http://genome.imb-jena.de/
                 Contact: gscj-submit@genome.imb-jena.de
            -------------- Project Information
                 Center project name: H3WZ, H4WZ, H5WZ, H6WZ, H7WZ, H8WZ, H9WZ
                 Center clone name: CTC-493P15, CTB-91N18, CTB-164D9,
                 CTB-169O5, CTB-65D4, CTC-271o23, CTC-367I24
            -------------- Summary Statistics
                 Sequencing vector: pUC18; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 319678 bases at least Q40
                 Consensus quality: 321154 bases at least Q30
                 Consensus quality: 321261 bases at least Q20
                 Quality coverage: 16.24x
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest.
            --------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .321261
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8 map 8p23-p22"
                     /clone="CTB-164D9, CTB-169o5, CTB-65D4, CTC-271o23,
                     CTC-367J24, complete sequence; and clones CTC-493P15,
                     CTB-91N18, partial sequence."
     source          1. .37559
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="CTB-91N18"
     source          1. .72370
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="CTC-493P15"
     source          1. .130843
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
```

```
                /clone="CTB-164D9"
     source     72368. .154799
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /clone="CTB-169O5"
     source     100525. .201673
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /clone="CTB-65D4"
     source     154796. .276418
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /clone="CTC-271o23"
     source     206159. .321261
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /clone="CTC-367I24"
     variation  1175. .1176
                /note="deleted in clone: CTC-493P15"
                /replace=""
     variation  3330
                /note="T substituted in clone: CTC-493P15"
                /replace="c"
     variation  3730
                /note="G substituted in clone: CTC-493P15"
                /replace="c"
     variation  5919
                /note="T substituted in clone: CTC-493P15"
                /replace="c"
     variation  6772
                /note="A substituted in clone: CTC-493P15"
                /replace="g"
     variation  6948
                /note="T substituted in clone: CTC-493P15"
                /replace="c"
     variation  7406
                /note="A substituted in clone: CTC-493P15"
                /replace="g"
     variation  7657
                /note="C substituted in clone: CTC-493P15"
                /replace="g"
     variation  7736. .7737
                /note="deleted in clone: CTC-493P15"
                /replace=""
     variation  8107
                /note="C substituted in clone: CTC-493P15"
                /replace="g"
     variation  8768
                /note="A substituted in clone: CTB-91N18 ,  CTB-164D9"
                /replace="g"
     variation  9096
                /note="T substituted in clone: CTC-493P15"
                /replace="c"
     variation  9325
                /note="T substituted in clone: CTC-493P15"
                /replace="c"
     variation  10007
                /note="G substituted in clone: CTC-493P15"
                /replace="t"
     variation  10175
                /note="A substituted in clone: CTC-493P15"
                /replace="t"
     variation  10361
                /note="A substituted in clone: CTB-164D9 ,  CTB-91N18"
                /replace="t"
     variation  10393
                /note="T substituted in clone: CTB-164D9 ,  CTB-91N18"
                /replace="a"
     variation  10435
                /note="G substituted in clone: CTC-493P15"
                /replace="a"
     variation  11300
                /note="T substituted in clone: CTC-493P15"
                /replace="c"
     variation  11834
                /note="T substituted in clone: CTC-493P15"
                /replace="a"
     variation  11944
                /note="A substituted in clone: CTB-164D9 ,  CTB-91N18"
                /replace="g"
     variation  12153
                /note="A substituted in clone: CTB-91N18 ,  CTB-164D9"
                /replace="c"
     variation  12234
                /note="A substituted in clone: CTC-493P15"
                /replace="t"
     variation  12321
                /note="A substituted in clone: CTC-493P15"
                /replace="g"
     variation  12631
                /note="C substituted in clone: CTC-493P15"
                /replace="t"
     variation  12739
                /note="A substituted in clone: CTC-493P15"
                /replace="g"
     variation  12809
                /note="G substituted in clone: CTC-493P15"
                /replace="a"
     variation  13386
                /note="T substituted in clone: CTC-493P15"
                /replace="c"
     variation  13973
                /note="A substituted in clone: CTC-493P15"
                /replace="g"
     variation  14023
                /note="T substituted in clone: CTC-493P15"
                /replace="c"
     variation  14037
                /note="A substituted in clone: CTC-493P15"
                /replace="t"

  Query Match             43.6%;  Score 159.2;  DB 9;  Length 321261;
  Best Local Similarity   72.4%;  Pred. No. 2.8e-33;
  Matches  220;  Conservative    0;  Mismatches   83;  Indels    1;  Gaps    1;

Qy         58 TTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCA 117
              | | || || |  ||  |||  | ||  | |  ||||| ||| ||||||||||||| ||||
Db      57470 TCTTTTTTTGTTTTTGTTTTCTTGTTGTTGTTGTTTTAAAGACAGGGTCTTGCTCTGTCA 57529

Qy        118 TCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCA 177
              || ||  |||| ||||| ||   |||||||||| |||| |  | || ||||||||| |||
Db      57530 TCCAGGCTGGAATGCAGTGGCACAATCATAGCTCACTGCAACCATGGATTCCTGAGCTCA 57589

Qy        178 AGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGC-TGCTACCATGCA 236
              ||  ||||| |||||||||| |||   ||||||| ||||| |  |||   | ||||  |
Db      57590 AGCAATCCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTAGAGCTGCACACCACCACACC 57649

Qy        237 CAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCT 296
               |||| |||||||||  |||||| ||||||||| ||||||| ||||| ||||||||||||
Db      57650 TAGCTAATTTTTAAAACTTTTTTATAGAGATGGGGTTGCCCCGGCTGATCTTGAACTCCT 57709

Qy        297 GGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCA 356
              || || | ||| | | ||||| |   ||||||||    ||  || |||||       |||
Db      57710 GGGCTCAAGTGTTTCCCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAACCA 57769

Qy        357 CCAC 360
              ||||
Db      57770 CCAC 57773


RESULT 9
AC023120/c
```

```
LOCUS       AC023120              188141 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-291C5, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC023120
VERSION     AC023120.3  GI:7230963
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188141)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188141)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 13, 2000 this sequence version replaced gi:7107585.

            ---------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0291C05
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 185645 bases at least Q40
            Consensus quality: 186465 bases at least Q30
            Consensus quality: 186911 bases at least Q20
            Insert size: 197000; agarose-fp
            Insert size: 187741; sum-of-contigs
            Quality coverage: 6.89 in Q20 bases; agarose-fp
            Quality coverage: 7.25 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7285: contig of 7285 bp in length
            *     7286     7385: gap of unknown length
            *     7386    29072: contig of 21687 bp in length
            *    29073    29172: gap of unknown length
            *    29173    80487: contig of 51315 bp in length
            *    80488    80587: gap of unknown length
            *    80588   132881: contig of 52294 bp in length
            *   132882   132981: gap of unknown length
            *   132982   188141: contig of 55160 bp in length.
FEATURES             Location/Qualifiers
     source          1. .188141
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="RP11-291C5"
     misc_feature    1. .7285
                     /note="assembly_name:Contig2
                     clone_end:SP6
                     vector_side:right"
     misc_feature    7386. .29072
                     /note="assembly_name:Contig3
                     clone_end:T7
                     vector_side:left"
     misc_feature    29173. .80487
                     /note="assembly_name:Contig4"
     misc_feature    80588. .132881
                     /note="assembly_name:Contig5"
     misc_feature    132982. .188141
                     /note="assembly_name:Contig6"
ORIGIN

  Query Match             42.1%;  Score 153.8;  DB 2;  Length 188141;
  Best Local Similarity   71.1%;  Pred. No. 8.7e-32;
  Matches  217;  Conservative    0;  Mismatches   87;  Indels    1;  Gaps    1;

Qy         53 TGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTC 112
              | | |||| || | ||  ||  |||  | ||  ||| ||||| |||| | |||||  |||
Db      34291 TTTTTTTTCTTTTCTTTTTTCTTTTTTTTTTTTTTTTTTTTTTTGAGACATGGTCTCACTC 34232

Qy        113 CGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGA 172
               |||  ||||  |||||||||| ||||  ||   | || |||| || |||||  || |||
Db      34231 TGTCTCCTAGGCTGGAGTGCAGTGGTGCGATTGCACCTCACTGCAGCCTTGACCTCTTGA 34172

Qy        173 GTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATG-TGCTGCTACC 231
              ||||||| |||||| || |||||||||| | |||| | ||| || |      |||||
Db      34171 GTTCAAGTGATCCTCCTACCTCAGCTTCCTAATTAGCCGGGACCATGGGCATGCCCTACC 34112

Qy        232 ATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAA 291
              |||| ||||| ||||||||   ||||||||  |||||| ||| ||||||||||||||||
Db      34111 ATGCCCAGCTAATTTTTAAGCGTTTTTTGTCAAGATGGGGTTTCCCAGGCTGGTCTTGAA 34052

Qy        292 CTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCA 351
              || |||| || | |  |||||||    ||| ||||||||    ||   | ||||| ||
Db      34051 CTGCTGGGCTCAAGCAATCCTCCCATCTTGGCCTCCCAAAGTGCTGGAATTACAGGTGTG 33992

Qy        352 CTCCA 356
                |||
Db      33991 AGCCA 33987


RESULT 10
AC133913
LOCUS       AC133913              192499 bp    DNA     linear   HTG 20-SEP-2002
DEFINITION  Homo sapiens chromosome 16 clone CTD-2312J12, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC133913
VERSION     AC133913.1  GI:23237947
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192499)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192499)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 747585
            Center clone name: CITB-H1_2312J12
            -------
            Summary Statistics
            Consensus quality: 183092 bases at least Q40
            Consensus quality: 187898 bases at least Q30
            Consensus quality: 189892 bases at least Q20
            Estimated insert size: 70000; agarose-fp estimation
```

```
            Estimated insert size: 191899; sum-of-contigs estimation
            Quality coverage:  19.36 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.06 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2449: contig of 2449 bp in length
            *     2450     2549: gap of unknown length
            *     2550    10498: contig of 7949 bp in length
            *    10499    10598: gap of unknown length
            *    10599    19985: contig of 9387 bp in length
            *    19986    20085: gap of unknown length
            *    20086    32520: contig of 12435 bp in length
            *    32521    32620: gap of unknown length
            *    32621    45536: contig of 12916 bp in length
            *    45537    45636: gap of unknown length
            *    45637    81480: contig of 35844 bp in length
            *    81481    81580: gap of unknown length
            *    81581   192499: contig of 110919 bp in length.
FEATURES             Location/Qualifiers
     source          1. .192499
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="CTD-2312J12"
                     /clone_lib="CalTech human BAC library D"
ORIGIN

  Query Match             42.0%;  Score 153.4;  DB 2;  Length 192499;
  Best Local Similarity   69.5%;  Pred. No. 1.1e-31;
  Matches  230;  Conservative    0;  Mismatches   86;  Indels   15;  Gaps    1;

Qy         48 CCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCT 107
              ||   ||  |||  ||||| |   |   |   | ||  ||| ||||| |||| |||||||
Db     115749 CCAAATGGTTTTCTTTCTTCTCTCTCTCTCTTTTTTTTTTTTTTTTTTTGAGACAGGGTCT 115808

Qy        108 TGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATT 167
              || || |||| ||||  ||||||| || ||||  |||||| || |||| | ||||||| |
Db     115809 TGTTCTGTCACCTAGGCTGGAGTGTAGTGGTGCCATCATAACTCACTGCAATCTTGAACT 115868

Qy        168 CCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGC 227
              |||||| ||||| |||| | ||||| |||| ||| | ||||||| ||||| | |||| ||
Db     115869 CCTGAGCTCAAGTGATCTTCCTGCCCCAGCCTCCTAAGTAGCTGGGACTACAGGTGCCGC 115928

Qy        228 TACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGGA----------------GT 272
              |||||||   ||| |||||||| |||||||| |  |   | ||                ||
Db     115929 CACCATGCCTGGCTAATTTTTAATTTTTTTTTTTTAAATAGAAATGGGGTCCTGCTATGT 115988

Qy        273 TGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGT 332
              ||| ||||||||||||||||||||||||| |  ||||||| ||    || ||||||||
Db     115989 TGCTCAGGCTGGTCTTGAACTCCTGGCCTCAAATGATCCTTCTTGTCTGGCCTCCCAAAG 116048

Qy        333 ATCTTAGACTACAGATGCACTCCACCACGCT 363
                |  || || || || |  |||||| |||
Db     116049 TGTTGGGATTATAGGTGTAAGCCACCATGCT 116079


RESULT 11
HSJ684F13
LOCUS       HSJ684F13             117230 bp    DNA     linear   PRI 31-OCT-2002
DEFINITION  Human DNA sequence from clone RP4-684F13 on chromosome 10, complete
            sequence.
ACCESSION   AL096706
VERSION     AL096706.10  GI:5869928
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 117230)
  AUTHORS   Bird,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 10, 1999 this sequence version replaced gi:5823996.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP4-684F13 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .117230
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704F13684"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP4-684F13"
                     /clone_lib="RPCI-4"
ORIGIN

  Query Match             41.8%;  Score 152.6;  DB 9;  Length 117230;
  Best Local Similarity   67.0%;  Pred. No. 1.9e-31;
  Matches  250;  Conservative    0;  Mismatches  114;  Indels    9;  Gaps    2;

Qy          2 TTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTA 61
              ||||| | ||  ||    | || || | |          || || |       | | ||||
Db       9792 TTGTGATTCTTTAATCTATGCTAAAGTATCCACTCAAAATTCATTTGATATTTTCTTTTT 9851

Qy         62 TTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTA 121
              ||||||  |||| |||  | ||  ||| ||||| ||||  | ||||  ||| |||| |||
Db       9852 TTCTTTCTATTTTTTTCTTTTTTTTTTTTTTTTGGAGATGGAGTCTCACTCTGTCACCTA 9911

Qy        122 GATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAG 181
              |  |||||||||| |||| ||||   ||| |||  |  ||      ||||||||||||| |
Db       9912 GGCTGGAGTGCAGTGGTGCAATCTCGGCTCACTACAACCTCCGCCTCCTGAGTTCAAGTG 9971

Qy        182 ATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGC 240
              || || |||||||||| ||||| || |||| || || | ||| ||||  |||||| | ||
Db       9972 ATTCTCCTGCCTCAGCCTCCCAAGTGGCTGGGATTACAGGTGCCTGCCTCCATGCCCGGC 10031

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGC--------CCAGGCTGGTCTTGAAC 292
```

```
              | ||||||  ||||||  |  |||   ||  ||||        ||||||||||||||||
Db      10032 TAATTTTTGTATTTTTAGTAGAGACGGGGTTTTGCCATGCTGGCCAGGCTGGTCTTGAAC 10091

Qy        293 TCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCAC 352
              |||||||| |||||||| ||||| ||| || |||||    ||  || ||||| ||
Db      10092 TCCTGGCCTCAGGTGATCCACCTGCCTTGGCCCCCCAAAGTGCTGGGATTACAGGTGTGA 10151

Qy        353 TCCACCACGCTTG 365
              |||    || ||
Db      10152 GCCATTGTGCCTG 10164


RESULT 12
AC084299
LOCUS       AC084299              168495 bp    DNA     linear   HTG 22-OCT-2000
DEFINITION  Homo sapiens chromosome 7 clone RP11-374D24, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AC084299
VERSION     AC084299.1  GI:10944492
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168495)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168495)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-OCT-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0374D24
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 154635 bases at least Q40
            Consensus quality: 159811 bases at least Q30
            Consensus quality: 162382 bases at least Q20
            Insert size: 170000; agarose-fp
            Insert size: 166695; sum-of-contigs
            Quality coverage: 4.03 in Q20 bases; agarose-fp
            Quality coverage: 4.16 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1283: contig of 1283 bp in length
            *     1284     1383: gap of unknown length
            *     1384     3418: contig of 2035 bp in length
            *     3419     3518: gap of unknown length
            *     3519     5659: contig of 2141 bp in length
            *     5660     5759: gap of unknown length
            *     5760     7665: contig of 1906 bp in length
            *     7666     7765: gap of unknown length
            *     7766    10469: contig of 2704 bp in length
            *    10470    10569: gap of unknown length
            *    10570    13355: contig of 2786 bp in length
            *    13356    13455: gap of unknown length
            *    13456    16432: contig of 2977 bp in length
            *    16433    16532: gap of unknown length
            *    16533    22300: contig of 5768 bp in length
            *    22301    22400: gap of unknown length
            *    22401    30627: contig of 8227 bp in length
            *    30628    30727: gap of unknown length
            *    30728    40948: contig of 10221 bp in length
            *    40949    41048: gap of unknown length
            *    41049    49251: contig of 8203 bp in length
            *    49252    49351: gap of unknown length
            *    49352    57626: contig of 8275 bp in length
            *    57627    57726: gap of unknown length
            *    57727    68382: contig of 10656 bp in length
            *    68383    68482: gap of unknown length
            *    68483    78948: contig of 10466 bp in length
            *    78949    79048: gap of unknown length
            *    79049    95469: contig of 16421 bp in length
            *    95470    95569: gap of unknown length
            *    95570   110400: contig of 14831 bp in length
            *   110401   110500: gap of unknown length
            *   110501   125387: contig of 14887 bp in length
            *   125388   125487: gap of unknown length
            *   125488   146498: contig of 21011 bp in length
            *   146499   146598: gap of unknown length
            *   146599   168495: contig of 21897 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168495
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP11-374D24"
     misc_feature    1. .1283
                     /note="assembly_name:Contig15"
     misc_feature    1384. .3418
                     /note="assembly_name:Contig16"
     misc_feature    3519. .5659
                     /note="assembly_name:Contig17"
     misc_feature    5760. .7665
                     /note="assembly_name:Contig18"
     misc_feature    7766. .10469
                     /note="assembly_name:Contig19"
     misc_feature    10570. .13355
                     /note="assembly_name:Contig20"
     misc_feature    13456. .16432
                     /note="assembly_name:Contig21"
     misc_feature    16533. .22300
                     /note="assembly_name:Contig22"
     misc_feature    22401. .30627
                     /note="assembly_name:Contig23"
     misc_feature    30728. .40948
                     /note="assembly_name:Contig24"
     misc_feature    41049. .49251
                     /note="assembly_name:Contig25"
     misc_feature    49352. .57626
                     /note="assembly_name:Contig26"
     misc_feature    57727. .68382
                     /note="assembly_name:Contig27"
     misc_feature    68483. .78948
                     /note="assembly_name:Contig28"
     misc_feature    79049. .95469
                     /note="assembly_name:Contig29"
     misc_feature    95570. .110400
                     /note="assembly_name:Contig30"
     misc_feature    110501. .125387
                     /note="assembly_name:Contig31"
     misc_feature    125488. .146498
                     /note="assembly_name:Contig32"
     misc_feature    146599. .168495
                     /note="assembly_name:Contig33"
ORIGIN
```

```
  Query Match             41.7%;  Score 152.2;  DB 2;  Length 168495;
  Best Local Similarity   70.8%;  Pred. No. 2.4e-31;
  Matches  216;  Conservative    0;  Mismatches   88;  Indels    1;  Gaps    1;

Qy         53 TGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTC 112
              | | |||  ||  |||  | | |||  | ||  ||| ||||| |||| | |||||  |||
Db     109052 TTTTTTTCTTTTCTTTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACATGGTCTCACTC 109111

Qy        113 CGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGA 172
              |||  ||||  |||||||||| ||||  ||   | || |||| || |||||  || |||
Db     109112 TGTCTCCTAGGCTGGAGTGCAGTGGTGCGATTGCACCTCACTGCAGCCTTGACCTCTTGA 109171

Qy        173 GTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATG-TGCTGCTACC 231
              ||||||| |||||| || ||||||||||| |  |||| | ||| ||  |       |||||
Db     109172 GTTCAAGTGATCCTCCTACCTCAGCTTCCTAATTAGCCGGGACCATGGGCATGCCCTACC 109231

Qy        232 ATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAA 291
              |||| ||||| ||||||||   ||||||||  |||||| ||| ||||||||||||||||
Db     109232 ATGCCCAGCTAATTTTTAAGCGTTTTTTGTCAAGATGGGGTTTCCCAGGCTGGTCTTGAA 109291

Qy        292 CTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCA 351
              || |||| || | |  |||||||    ||| ||||||||     ||    | ||||| ||
Db     109292 CTGCTGGGCTCAAGCAATCCTCCCATCTTGGCCTCCCAAAGTGCTGGAATTACAGGTGTG 109351

Qy        352 CTCCA 356
                |||
Db     109352 AGCCA 109356
```

```
RESULT 13
AC136144/c
LOCUS       AC136144               162216 bp    DNA     linear   HTG 01-OCT-2004
DEFINITION  Pan troglodytes clone rp43-47m18, WORKING DRAFT SEQUENCE, 5 ordered
            pieces.
ACCESSION   AC136144
VERSION     AC136144.14  GI:53373921
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 162216)
  AUTHORS   Lau,C. and Roe,B.A.
  TITLE     Pan troglodytes BAC Clone rp43-47m18
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162216)
  AUTHORS   Lau,C. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 162216)
  AUTHORS   Lau,C. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Oct 1, 2004 this sequence version replaced gi:42716103.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1     7544: contig of 7544 bp in length
            *    7545     7644: gap of unknown length
            *    7645    23317: contig of 15673 bp in length
            *   23318    23417: gap of unknown length
            *   23418   102290: contig of 78873 bp in length
            *  102291   102390: gap of unknown length
            *  102391   132494: contig of 30104 bp in length
            *  132495   132594: gap of unknown length
            *  132595   162216: contig of 29622 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162216
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="rp43-47m18"
                     /clone_lib="RPCI - 43 Male Chimpanzee BAC Library"
ORIGIN
```

```
  Query Match             41.6%;  Score 152;  DB 2;  Length 162216;
  Best Local Similarity   69.6%;  Pred. No. 2.7e-31;
  Matches  224;  Conservative    0;  Mismatches   90;  Indels    8;  Gaps    1;

Qy         52 GTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCT 111
              | || |||| || ||||| |||||||  | |   ||  ||||| ||||  | |||| |||
Db      38851 GGGTTTTTTTTTGTTTTGTTTTGTTTGTTTTGTTTTGTTTTTTTGAGACGGTGTCTCGCT 38792

Qy        112 CCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTG 171
              | ||| | ||  |||||||||| || | ||||  | ||| ||| |||||      |||||
Db      38791 GCATCACCCAGGCTGGAGTGCAGTGGCGCAATCTCACCTTCCTGCAGTCTCCGCCTCCTG 38732

Qy        172 AGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACC 231
              |||||||| ||| || |||||||||| ||||  |||||||  |||||  | | ||    |||
Db      38731 AGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCGCACACC 38672

Qy        232 ATGCACAGCTGATTTTTAAATTTTTTTTGTAGA--------GATGGAGTTGCCCAGGCTG 283
              | || ||||| ||||||  ||||||  |  |||          ||   |||| ||||||||
Db      38671 ACGCCCAGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCATCATGTTGACCAGGCTG 38612

Qy        284 GTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTA 343
              ||||  |||||||| | | ||||||||| || || |   |||||||  |  ||  || ||
Db      38611 GTCTCAAACTCCTGACTTCAGGTGATCCACCCGCCTCAGCCTCCCAACTTGCTGGGATTA 38552

Qy        344 CAGATGCACTCCACCACGCTTG 365
              |||  |    || ||||||  ||
Db      38551 CAGCCGTGAGCCGCCACGCCTG 38530
```

```
RESULT 14
AC010397
LOCUS       AC010397               151408 bp    DNA     linear   HTG 06-MAY-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2122L23, WORKING DRAFT
            SEQUENCE, 9 ordered pieces.
ACCESSION   AC010397
VERSION     AC010397.5  GI:7711436
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151408)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151408)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On May 6, 2000 this sequence version replaced gi:7710717.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
```

```
           -------
           Project Information
           Center Project Name: 674684
           Center clone name: CITB-H1_2122L23
           -------
           Summary Statistics
           Consensus quality: 137346 bases at least Q40
           Consensus quality: 146291 bases at least Q30
           Consensus quality: 148789 bases at least Q20
           Estimated insert size: 144000; pulse field gel estimation
           Estimated insert size: 150608; sum-of-contigs estimation
           Quality coverage:  6.06 in Q20 bases; pulse field gel estimation
           Quality coverage:  5.78 in Q20 bases; sum-of-contigs estimation.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 9 contigs. Gaps between the contigs
           * are represented as runs of N. The order of the pieces
           * is believed to be correct as given, however the sizes
           * of the gaps between them are based on estimates that have
           * provided by the submittor.
           * This sequence will be replaced
           * by the finished sequence as soon as it is available and
           * the accession number will be preserved.
           *        1    52908: contig of 52908 bp in length
           *    52909    53008: gap of unknown length
           *    53009    74920: contig of 21912 bp in length
           *    74921    75020: gap of unknown length
           *    75021    78508: contig of 3488 bp in length
           *    78509    78608: gap of unknown length
           *    78609    84262: contig of 5654 bp in length
           *    84263    84362: gap of unknown length
           *    84363   109884: contig of 25522 bp in length
           *   109885   109984: gap of unknown length
           *   109985   117979: contig of 7995 bp in length
           *   117980   118079: gap of unknown length
           *   118080   120817: contig of 2738 bp in length
           *   120818   120917: gap of unknown length
           *   120918   140301: contig of 19384 bp in length
           *   140302   140401: gap of unknown length
           *   140402   151408: contig of 11007 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151408
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2122L23"
                     /clone_lib="CalTech human BAC library D"
ORIGIN

  Query Match             41.5%;  Score 151.4;  DB 2;  Length 151408;
  Best Local Similarity   71.8%;  Pred. No. 4e-31;
  Matches  229;  Conservative    0;  Mismatches   81;  Indels    9;  Gaps    2;

Qy          53 TGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTC 112
               ||   | | |||||||  ||| |||    ||  ||  ||| |||||| |  |||||||||
Db      104495 TGATGTGTTTTCTTTTTTTTTTTTTTGTTTTTTGTTTTGTAGAGACAATGTCTTGCTC 104554

Qy         113 CGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGA 172
                |||| | ||  |||||||||| ||||| ||||||||| ||||| | ||  ||  ||||
Db      104555 TGTCACCCAGGCTGGAGTGCAGTGGTGTGATCATAGCTCACTGTGGCCTGAAACCCCTGG 104614

Qy         173 GTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGC-TGCTACC 231
               | ||||| ||| || |||||||||| |||   |||||||||||| | ||||  || |||
Db      104615 GCTCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGAGACTACAGGTGCACGCCACC 104674

Qy         232 ATGCACAGCTGATTTTTAAATTTTTTTTGTAGA--------GATGGAGTTGCCCAGGCTG 283
               |||| ||||| | |||| || ||||| |  |||          |  | ||||||||||||
Db      104675 ATGCTCAGCTAACTTTTTAAATTTTTGTAGAGACTGGGTCTCACTGTGTTGCCCAGGCTA 104734

Qy         284 GTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTA 343
               |||| |||||| |||||| | |||||||||||||| ||||| |||||||  | |    | ||
Db      104735 GTCTCGAACTCTTGGCCTCAAGTGATCCTCCTGCCTTGACATCCCAAGGCTTTGGAATTA 104794

Qy         344 CAGATGCACTCCACCACGC 362
               |||        |||||| ||
Db      104795 CAGGCATGAGCCACCATGC 104813


RESULT 15
AC146346
LOCUS       AC146346              155046 bp    DNA     linear   HTG 12-AUG-2003
DEFINITION  Pan troglodytes clone rp43-26h17, WORKING DRAFT SEQUENCE, 46
            unordered pieces.
ACCESSION   AC146346
VERSION     AC146346.1  GI:33589913
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 155046)
  AUTHORS   Li,J. and Roe,B.A.
  TITLE     Pan troglodytes BAC Clone rp43-26h17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155046)
  AUTHORS   Li,J. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     ---------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 46 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2373: contig of 2373 bp in length
            *    2374     2473: gap of unknown length
            *    2474     4532: contig of 2059 bp in length
            *    4533     4632: gap of unknown length
            *    4633     6863: contig of 2231 bp in length
            *    6864     6963: gap of unknown length
            *    6964     9623: contig of 2660 bp in length
            *    9624     9723: gap of unknown length
            *    9724    11788: contig of 2065 bp in length
            *   11789    11888: gap of unknown length
            *   11889    14353: contig of 2465 bp in length
            *   14354    14453: gap of unknown length
            *   14454    16590: contig of 2137 bp in length
            *   16591    16690: gap of unknown length
            *   16691    18694: contig of 2004 bp in length
            *   18695    18794: gap of unknown length
            *   18795    21483: contig of 2689 bp in length
            *   21484    21583: gap of unknown length
            *   21584    24281: contig of 2698 bp in length
            *   24282    24381: gap of unknown length
            *   24382    26378: contig of 1997 bp in length
            *   26379    26478: gap of unknown length
            *   26479    28924: contig of 2446 bp in length
            *   28925    29024: gap of unknown length
            *   29025    31499: contig of 2475 bp in length
            *   31500    31599: gap of unknown length
            *   31600    33801: contig of 2202 bp in length
            *   33802    33901: gap of unknown length
            *   33902    36444: contig of 2543 bp in length
            *   36445    36544: gap of unknown length
            *   36545    38793: contig of 2249 bp in length
            *   38794    38893: gap of unknown length
            *   38894    40958: contig of 2065 bp in length
```

```
            *    40959    41058: gap of unknown length
            *    41059    43163: contig of 2105 bp in length
            *    43164    43263: gap of unknown length
            *    43264    46061: contig of 2798 bp in length
            *    46062    46161: gap of unknown length
            *    46162    48520: contig of 2359 bp in length
            *    48521    48620: gap of unknown length
            *    48621    51081: contig of 2461 bp in length
            *    51082    51181: gap of unknown length
            *    51182    53575: contig of 2394 bp in length
            *    53576    53675: gap of unknown length
            *    53676    55755: contig of 2080 bp in length
            *    55756    55855: gap of unknown length
            *    55856    58155: contig of 2300 bp in length
            *    58156    58255: gap of unknown length
            *    58256    60930: contig of 2675 bp in length
            *    60931    61030: gap of unknown length
            *    61031    63093: contig of 2063 bp in length
            *    63094    63193: gap of unknown length
            *    63194    65639: contig of 2446 bp in length
            *    65640    65739: gap of unknown length
            *    65740    68428: contig of 2689 bp in length
            *    68429    68528: gap of unknown length
            *    68529    70796: contig of 2268 bp in length
            *    70797    70896: gap of unknown length
            *    70897    74328: contig of 3432 bp in length
            *    74329    74428: gap of unknown length
            *    74429    77647: contig of 3219 bp in length
            *    77648    77747: gap of unknown length
            *    77748    81290: contig of 3543 bp in length
            *    81291    81390: gap of unknown length
            *    81391    84510: contig of 3120 bp in length
            *    84511    84610: gap of unknown length
            *    84611    86858: contig of 2248 bp in length
            *    86859    86958: gap of unknown length
            *    86959    89999: contig of 3041 bp in length
            *    90000    90099: gap of unknown length
            *    90100    93701: contig of 3602 bp in length
            *    93702    93801: gap of unknown length
            *    93802    98578: contig of 4777 bp in length
            *    98579    98678: gap of unknown length
            *    98679   103341: contig of 4663 bp in length
            *   103342   103441: gap of unknown length
            *   103442   107766: contig of 4325 bp in length
            *   107767   107866: gap of unknown length
            *   107867   112674: contig of 4808 bp in length
            *   112675   112774: gap of unknown length
            *   112775   119992: contig of 7218 bp in length
            *   119993   120092: gap of unknown length
            *   120093   126312: contig of 6220 bp in length
            *   126313   126412: gap of unknown length
            *   126413   131814: contig of 5402 bp in length
            *   131815   131914: gap of unknown length
            *   131915   138652: contig of 6738 bp in length
            *   138653   138752: gap of unknown length
            *   138753   145691: contig of 6939 bp in length
            *   145692   145791: gap of unknown length
            *   145792   155046: contig of 9255 bp in length.
FEATURES             Location/Qualifiers
     source          1. .155046
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="rp43-26h17"
                     /clone_lib="RPCI - 43 Male Chimpanzee BAC Library"
ORIGIN

  Query Match           41.5%;  Score 151.4;  DB 2;  Length 155046;
  Best Local Similarity   69.3%;  Pred. No. 4e-31;
  Matches  228;  Conservative    0;  Mismatches   86;  Indels   15;  Gaps    1;

Qy         52 GTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCT 111
              || |  ||| |||| |   | | | |    ||  ||| | ||| |||| |||||||||| |

Db     135665 GTTTTCTTTCTTCTCTCTCTCTCTCTCTTTTTTTTTTCTTTTGAGACAGGGTCTTGTT 135724

Qy         112 CCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTG 171
               | |||| ||||  ||||||| || ||||  |||||| || |||| | | ||||| |||||
Db     135725 CTGTCACCTAGGCTGGAGTGTAGTGGTGCCATCATAACTCACTGCAATATTGAACTCCTG 135784

Qy         172 AGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACC 231
               || ||||| |||| | ||||| |||| ||| | ||||||| ||||| | |||| || |||
Db     135785 AGCTCAAGTGATCTTCCTGCCCCAGCCTCCTAAGTAGCTGGGACTACAGGTGCCGCCACC 135844

Qy         232 ATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGGA---------------GTTGCC 276
               ||||   ||| |||||||| |||||||| |  |   | |               |||||
Db     135845 ATGCCTGGCTAATTTTTAATTTTTTTTTTTTTAAATAGAAATGGGGTCCTGCTATGTTGCT 135904

Qy         277 CAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCT 336
               |||||||||||||||||||||||| |  ||||||| ||    || ||||||||     |
Db     135905 CAGGCTGGTCTTGAACTCCTGGCCTCAAATGATCCTTCTTGTCTGGCCTCCCAAAGTGTT 135964

Qy         337 TAGACTACAGATGCACTCCACCACGCTTG 365
                  || || || || |  |||||| ||| |
Db     135965 GGGATTATAGGTGTAAGCCACCATGCTCG 135993

Search completed: August 30, 2005, 04:38:40
Job time : 2076 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 02:32:53 ; Search time 369 Seconds
(without alignments)
5855.571 Million cell updates/sec

Title: US-08-731-499-9_COPY_10001_10365
Perfect score: 365
Sequence: 1 TTTGTGGTCCTCCAAGGCTT.....GATGCACTCCACCACGCTTG 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	365	100.0	14906	4	AAK81093	Aak81093 Human imm
	2	365	100.0	36022	11	ACN44986	Acn44986 Human gen
	3	325.8	89.3	10282	2	AAV09023	Aav09023 Homo sapi
	4	147.6	40.4	310268	13	ABD32548	Abd32548 Human can
c	5	147.2	40.3	50335	9	AAD58280	Aad58280 Human tum
c	6	147.2	40.3	226475	9	AAD58279	Aad58279 Human tum
	7	146.6	40.2	1385	4	AAK70207	Aak70207 Human imm
c	8	146.6	40.2	1385	4	AAK78729	Aak78729 Human imm
	9	145.8	39.9	110000	13	ABD32780_2	Continuation (3 of
	10	145.2	39.8	3019	12	ADQ64936	Adq64936 Novel hum
	11	145	39.7	10735	8	ABZ58995	Abz58995 Human onc
c	12	143.6	39.3	11216	4	AAK88992	Aak88992 Human dig
c	13	143.6	39.3	11216	5	AAS39593	Aas39593 Genomic s
c	14	143.6	39.3	11216	9	ADB32553	Adb32553 Human nov
c	15	142.8	39.1	14540	4	AAK72853	Aak72853 Human imm
	16	142.6	39.1	57502	12	ADQ97092	Adq97092 Human can
	17	142.6	39.1	78082	12	ADQ97968	Adq97968 Human can
	18	142.6	39.1	155937	12	ADQ19389	Adq19389 Human sof
c	19	142.4	39.0	1068	4	AAK78726	Aak78726 Human imm
	20	142	38.9	14327	4	AAK79116	Aak79116 Human imm
	21	142	38.9	14327	8	ABZ74029	Abz74029 Secreted
	22	142	38.9	14327	8	ADA98634	Ada98634 Human sec
	23	142	38.9	14327	10	ABT16938	Abt16938 Human sec
	24	142	38.9	14327	10	ABZ67608	Abz67608 Human sec
c	25	141.8	38.8	17596	4	AAK72852	Aak72852 Human imm
	26	140.8	38.6	84607	2	AAX90847	Aax90847 Human PAC
	27	140.4	38.5	22930	4	AAK78545	Aak78545 Human imm
	28	140.4	38.5	138363	13	ABD32624	Abd32624 Human can
	29	140.2	38.4	27675	4	AAK85837	Aak85837 Human imm
c	30	140.2	38.4	169659	12	ADQ59434	Adq59434 Human can
	31	140	38.4	2757	12	ADQ63687	Adq63687 Novel hum
	32	140	38.4	14346	4	AAS32551	Aas32551 Human gen
c	33	140	38.4	20303	2	AAT71699	Aat71699 Human deo
c	34	140	38.4	26764	2	AAT71696	Aat71696 Human deo
	35	140	38.4	26865	12	ADM97421	Adm97421 Prostate
c	36	140	38.4	34739	11	ACN45078	Acn45078 Human gen
	37	140	38.4	171158	12	ADQ97894	Adq97894 Human can
c	38	139.8	38.3	5230	5	AAF73808	Aaf73808 Partial h
c	39	139.8	38.3	5230	10	ACC49356	Acc49356 Human 5HT
	40	139.6	38.2	26928	5	ABA82620	Aba82620 Human HBM
	41	139.6	38.2	26928	6	ABK22779	Abk22779 Human hig
	42	139.6	38.2	26928	8	ACC45361	Acc45361 Human HBM
	43	139.6	38.2	26928	10	ADB98061	Adb98061 HBM-relat
	44	139.6	38.2	26928	10	ADE82430	Ade82430 Human DNA
	45	139.6	38.2	26928	13	ADR16924	Adr16924 BAC clone

ALIGNMENTS

RESULT 1
AAK81093
ID AAK81093 standard; DNA; 14906 BP.
XX
AC AAK81093;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35905.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 35905; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 14906 BP; 3959 A; 3088 C; 3333 G; 4526 T; 0 U; 0 Other;

```
  Query Match             100.0%;  Score 365;  DB 4;  Length 14906;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-88;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6404 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 6463

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6464 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 6523

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6524 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 6583

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6584 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 6643

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6644 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 6703

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6704 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 6763

Qy        361 GCTTG 365
              |||||
Db       6764 GCTTG 6768
```

RESULT 2
ACN44986
ID ACN44986 standard; DNA; 36022 BP.
XX
AC ACN44986;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG37127.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 1708; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 36022 BP; 9645 A; 7326 C; 8145 G; 10906 T; 0 U; 0 Other;

```
  Query Match             100.0%;  Score 365;  DB 11;  Length 36022;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-88;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17531 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 17590

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17591 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 17650

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17651 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 17710

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17711 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 17770

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17771 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 17830

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17831 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 17890

Qy        361 GCTTG 365
              |||||
Db      17891 GCTTG 17895
```

RESULT 3
AAV09023
ID AAV09023 standard; DNA; 10282 BP.
XX
AC AAV09023;
XX
DT 21-JUL-1998 (first entry)
XX
DE Homo sapiens 20q13 amplicon ZABC-1 genomic sequence.
XX
KW 20q13 amplicon; chromosome 20; tumour; detection; ZABC-1 gene;
KW chromosomal abnormalities; probe; gene therapy; antisense inhibition;
KW treatment; age-related macular degeneration; retinitis pigmentation;
KW Leber's congenital amaurosis; zinc finger amplified in breast cancer; ds.
XX
OS Homo sapiens.
XX
PN WO9802539-A1.
XX
PD 22-JAN-1998.
XX
PF 15-JUL-1997; 97WO-US012343.
XX
PR 15-JUL-1996; 96US-00680395.
PR 16-OCT-1996; 96US-00731499.
PR 17-JAN-1997; 97US-00785532.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Gray JW, Collins CC, Hwang S, Godfrey T, Kowbel D, Rommens J;

```
XX
DR   WPI; 1998-110587/10.
DR   P-PSDB; AAW23975.
XX
PT   New sequences from the 20q13 amplicon - used for detecting chromosomal
PT   abnormalities, particularly tumours, and for developing products for
PT   treating diseases.
XX
PS   Claim 1; Page 64-67; 91pp; English.
XX
CC   The sequence is that of the genomic sequence of ZABC-1 (zinc finger
CC   amplified in breast cancer). It maps to the core of the 20q13.2 amplicon
CC   and is overexpressed in primary tumours and breast cancer cell lines
CC   having 20q13.2 amplification. The exact coding region for the genomic
CC   sequence is not given. The sequence can be used as a probe for the
CC   detection of chromosomal abnormalities at 20q13. It and other sequences
CC   isolated from the 20q13 amplicon are consistently amplified in primary
CC   tumours. These sequences are useful as probes or as probe targets for
CC   monitoring the relative copy number of corresponding sequences from a
CC   biological sample such as tumour cells. The sequences can also be used in
CC   therapeutic applications for modulating the expression of the endogenous
CC   gene or the activity of the gene product. Examples of therapeutic
CC   approaches include antisense inhibition of gene expression, gene therapy,
CC   and monoclonal antibodies that specifically bind the gene products. The
CC   products can also be used in the treatment of other diseases, e.g. age-
CC   related macular degeneration, Leber's congenital amaurosis and retinitis
CC   pigmentation
XX
SQ   Sequence 10282 BP; 2820 A; 2222 C; 2191 G; 3047 T; 0 U; 2 Other;

  Query Match             89.3%;  Score 325.8;  DB 2;  Length 10282;
  Best Local Similarity   95.1%;  Pred. No. 1.2e-77;
  Matches  347;  Conservative    0;  Mismatches   17;  Indels    1;  Gaps    1;

Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||| |   ||||||||||||||| | ||||||
Db       9919 TTTGTGGTCCTCCAAGGCTTACTTAACTCTCTCCCTTTAACTCTTAACCCTCTCTATTTT 9978

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              |||||||||||||||||||||||||||||||||||||||  ||||||||||||||||||
Db       9979 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAACCGTCTTGCTCCGTCATCT 10038

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              |||||||||||||||||||||||||||||||||||| |||||||||||| |||||||||
Db      10039 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTACTCTTGAATTCCTCAGTTCAAGA 10098

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              |||||||||||||||||||||||||| |||||||||||||| || ||||||| ||||||
Db      10099 GATCCTTCTGCCTCAGCTTCCCAGGTACCTGAGACTATATGTCCTCCTACCATCCACAGC 10158

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              | |||||||||| |||||||||| |||||||||||||||||||| ||||||||||||||||||
Db      10159 TCATTTTTAAA-TTTTTTTGTACAGATGGAGTTGCCCAGGCTCGTCTTGAACTCCTGGCC 10217

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10218 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 10277

Qy        361 GCTTG 365
              |||||
Db      10278 GCTTG 10282


RESULT 4
ABD32548
ID   ABD32548 standard; DNA; 310268 BP.
XX
AC   ABD32548;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD14-033.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 24; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 310268 BP; 87522 A; 60932 C; 62901 G; 98913 T; 0 U; 0 Other;

  Query Match             40.4%;  Score 147.6;  DB 13;  Length 310268;
  Best Local Similarity   65.7%;  Pred. No. 4.1e-29;
  Matches  234;  Conservative    0;  Mismatches  114;  Indels    8;  Gaps    1;

Qy         18 CTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTT 77
              ||  | || |  | |||    |||||  |  |   |   |    |  |  |  ||| |||
Db      77830 CTCTCATAGCATTATGGAAACAACTCGAATACAACTACGTCAATTATTCATCTTTTTTTT 77889
```

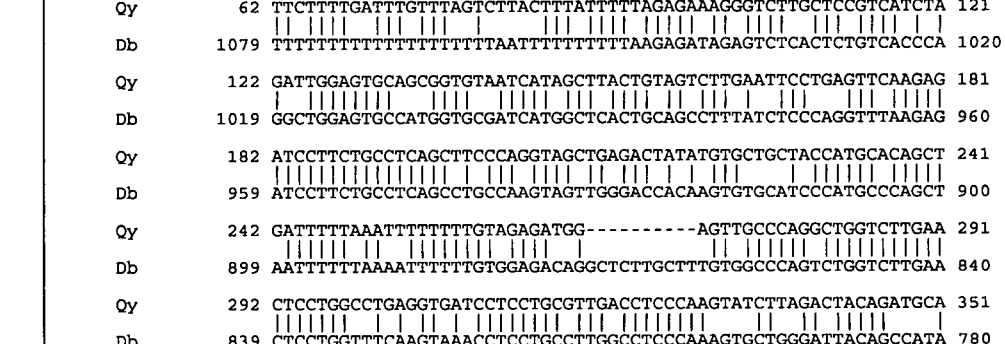
```
Qy         78 AGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGG 137
              | ||  ||| ||||| |||| || |||   |||| ||| | ||  |||||||||| ||
Db      77890 TTTTTTTTTTTTTTTTTTGAGACAGAGTCCCACTCCATCAACCAGGCTGGAGTGCAGTGG 77949

Qy        138 TGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGC 197
              || | || | ||| |||| | ||||    ||||| | ||||| ||| ||  |||||||||
Db      77950 TGCAGTCTTGGCTCACTGCATTCTTTGCATCCTGGGCTCAAGTGATTCTCGTGCCTCAGC 78009

Qy        198 TTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCTGATTTTTAAATTTTTT 257
              ||||| |||||||||| || | ||||     ||||| | ||| | ||||||  ||||||
Db      78010 CTCCCAAGTAGCTGAGATTACAGGTGCCCACACCATTCTCAGATAATTTTTGTATTTTTA 78069

Qy        258 TTGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGAT 309
              |  |||   ||         |||| |||||||||||| ||||||||| ||| | |||||
Db      78070 CTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAAGTGAT 78129

Qy        310 CCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
              || ||||  ||| ||||||||     ||  || ||||| |||   ||||   || ||
Db      78130 CCACCTGTCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGCGAACCACTGTGCCTG 78185


RESULT 5
AAD58280/c
ID   AAD58280 standard; DNA; 50335 BP.
XX
AC   AAD58280;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human tumour suppressor gene, Lmt intron 1 DNA.
XX
KW   Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003066869-A1.
XX
PD   14-AUG-2003.
XX
PF   07-FEB-2003; 2003WO-AU000126.
XX
PR   07-FEB-2002; 2002AU-00000371.
XX
PA   (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI   Cook WD,  Mccaw BJ;
XX
DR   WPI; 2003-646311/61.
XX
PT   New nucleic acid molecule, useful for screening a subject for the
PT   presence of an aberration in a gene encoding an LMT.
XX
PS   Claim 10; Page 299-314; 373pp; English.
XX
CC   The invention relates to novel tumour suppressor gene, referred to as
CC   Lmt. The invention also relates to the field of cancer therapy and cancer
CC   diagnostics. The nucleic acid molecule is useful for screening a subject
CC   for the presence of an aberration in a gene encoding an LMT. The present
CC   sequence is human Lmt intron 1 DNA
XX
SQ   Sequence 50335 BP; 15699 A; 9846 C; 9072 G; 14424 T; 0 U; 1294 Other;

  Query Match             40.3%;  Score 147.2;  DB 9;  Length 50335;
  Best Local Similarity   74.5%;  Pred. No. 3.1e-29;
  Matches  213;  Conservative    0;  Mismatches   68;  Indels    5;  Gaps    2;

Qy         75 TTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAG 134
              | |  | ||  ||| ||||| |||| | |||||  ||| |||||| ||  || |||||||
Db      35888 TCTTTTTTTTTTTTTTTTTTTGAGACAAGGTCTCACTCTGTCATCCAGGCTGAAGTGCAG 35829

Qy        135 CGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTC 194
              ||  | ||||| ||| | ||||| || ||  ||||||| ||| | |||||| |  ||||
Db      35828 TGGCATGATCATGGCTCAGTGTAGCCTCGACCTCCTGAGCTCAGGTGATCCTCCCACCTC 35769

Qy        195 AGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGATTTTTAAATT 253
              ||| |||||||||||||| |||||||  | |  ||||||||||  |||| ||||||||||
Db      35768 AGCCTCCCAGGTAGCTGGGACTATAGGAGTGTGCTACCATGCCTAGCTAATTTTTAAATT 35709

Qy        254 TTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTC 313
              ||| | |      |  ||||||||||||||||| |||||||||| || |||||||| |
Db      35708 CTTTGTTT----GTTATGTTGCCCAGGCTGGTCTCGAACTCCTGGGCTCAGGTGATCCAC 35653

Qy        314 CTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCA 359
              | || ||| |||||||     || ||| |||||| |    ||||||
Db      35652 CCGCCTTGGCCTCCCACAGTGCTGAGATTACAGACGTGAGCCACCA 35607


RESULT 6
AAD58279/c
ID   AAD58279 standard; DNA; 226475 BP.
XX
AC   AAD58279;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human tumour suppressor gene, Lmt reverse complement DNA.
XX
KW   Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003066869-A1.
XX
PD   14-AUG-2003.
XX
PF   07-FEB-2003; 2003WO-AU000126.
XX
PR   07-FEB-2002; 2002AU-00000371.
XX
PA   (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.
XX
PI   Cook WD,  Mccaw BJ;
XX
DR   WPI; 2003-646311/61.
XX
PT   New nucleic acid molecule, useful for screening a subject for the
PT   presence of an aberration in a gene encoding an LMT.
XX
PS   Claim 10; Page 233-299; 373pp; English.
XX
CC   The invention relates to novel tumour suppressor gene, referred to as
CC   Lmt. The invention also relates to the field of cancer therapy and cancer
CC   diagnostics. The nucleic acid molecule is useful for screening a subject
CC   for the presence of an aberration in a gene encoding an LMT. The present
CC   sequence is human Lmt reverse complement DNA
XX
SQ   Sequence 226475 BP; 61024 A; 41761 C; 40916 G; 57494 T; 0 U; 25280 Other;

  Query Match             40.3%;  Score 147.2;  DB 9;  Length 226475;
  Best Local Similarity   74.5%;  Pred. No. 4.8e-29;
  Matches  213;  Conservative    0;  Mismatches   68;  Indels    5;  Gaps    2;

Qy         75 TTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAG 134
              | |  | ||  ||| ||||| |||| | |||||  ||| |||||| ||  || |||||||
Db     211265 TCTTTTTTTTTTTTTTTTTTTTGAGACAAGGTCTCACTCTGTCATCCAGGCTGAAGTGCAG 211206

Qy        135 CGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTC 194
              ||  | ||||| ||| | ||||| || ||  ||||||| ||| | |||||| |  ||||
Db     211205 TGGCATGATCATGGCTCAGTGTAGCCTCGACCTCCTGAGCTCAGGTGATCCTCCCACCTC 211146

Qy        195 AGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGATTTTTAAATT 253
              ||| |||||||||||||| |||||||  | |  ||||||||||  |||| ||||||||||
Db     211145 AGCCTCCCAGGTAGCTGGGACTATAGGAGTGTGCTACCATGCCTAGCTAATTTTTAAATT 211086

Qy        254 TTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTC 313
              ||| | |      |  ||||||||||||||||| |||||||||| || |||||||| |
```

```
Db     211085 CTTTGTTT----GTTATGTTGCCCAGGCTGGTCTCGAACTCCTGGGCTCAGGTGATCCAC 211030

Qy        314 CTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCA 359
              | || ||| ||||||| || ||| |||||| |  ||||||
Db     211029 CCGCCTTGGCCTCCCACAGTGCTGAGATTACAGACGTGAGCCACCA 210984


RESULT 7
AAK70207
ID   AAK70207 standard; DNA; 1385 BP.
XX
AC   AAK70207;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25019.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
```

```
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 25019; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 1385 BP; 349 A; 289 C; 278 G; 469 T; 0 U; 0 Other;

  Query Match             40.2%;  Score 146.6;  DB 4;  Length 1385;
  Best Local Similarity   69.8%;  Pred. No. 1.6e-29;
  Matches  217;  Conservative    0;  Mismatches   84;  Indels   10;  Gaps    1;

Qy         62 TTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTA 121
              || ||||  ||| |||  |     ||| |||| ||||| || ||||  ||| |||| | |
Db        307 TTTTTTTTTTTTTTTTTTTTAATTTTTTTTTTTAAGAGATAGAGTCTCACTCTGTCACCCA 366

Qy        122 GATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAG 181
              |  ||||||||   ||||  ||||| ||| |||| || ||| |  |||   ||| |||||
Db        367 GGCTGGAGTGCCATGGTGCGATCATGGCTCACTGCAGCCTTTATCTCCCAGGTTTAAGAG 426

Qy        182 ATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCT 241
              ||||||||||||||||  | ||| |||| || ||| | | |||    | |||||| |||||
Db        427 ATCCTTCTGCCTCAGCCTGCCAAGTAGTTGGGACCACAAGTGTGCATCCCATGCCCAGCT 486

Qy        242 GATTTTTAAATTTTTTTTGTAGAGATGG----------AGTTGCCCAGGCTGGTCTTGAA 291
              ||||||  ||  ||||||||  ||||  |           ||  ||||||  ||||||||||
Db        487 AATTTTTTAAAATTTTTTGTGGAGACAGGCTCTTGCTTTGTGGCCCAGTCTGGTCTTGAA 546

Qy        292 CTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCA 351
              |||||||  | | ||  | ||||||||  ||| ||||||||     ||  ||  |||||    |
Db        547 CTCCTGGTTTCAAGTAAACCTCCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGCCATA 606

Qy        352 CTCCACCACGC 362
                  |||||| ||
Db        607 AGCCACCATGC 617


RESULT 8
AAK78729/c
ID   AAK78729 standard; DNA; 1385 BP.
XX
AC   AAK78729;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33541.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
```

PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 33541; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 1385 BP; 469 A; 278 C; 289 G; 349 T; 0 U; 0 Other;

Query Match 40.2%; Score 146.6; DB 4; Length 1385;
Best Local Similarity 69.8%; Pred. No. 1.6e-29;
Matches 217; Conservative 0; Mismatches 84; Indels 10; Gaps 1;



```
Qy        352 CTCCACCACGC 362
              |||||| ||
Db        779 AGCCACCATGC 769


RESULT 9
ABD32780_2
Continuation (3 of 5) of ABD32780 from base 200001 (Human cancer-associated genomic DNA
WP Sequence split into 5 fragments  LOCUS ABD32780 Accession Abd32780
WP      Fragment Name       Begin      End
WP      ABD32780_0              1    110000
WP      ABD32780_1         100001    210000
WP      ABD32780_2         200001    310000
WP      ABD32780_3         300001    410000
WP      ABD32780_4         400001    430442

  Query Match             39.9%;  Score 145.8;  DB 13;  Length 110000;
  Best Local Similarity   67.3%;  Pred. No. 9.2e-29;
  Matches  239;  Conservative    0;  Mismatches  107;  Indels    9;  Gaps    2;

Qy         19 TTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTA 78
              | || ||  ||| | | | | | | |       | |   |||  ||  |||  ||| |||
Db      21592 TCACCTATTTCTTTTGATGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTT 21651

Qy         79 GTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGT 138
              | ||| ||||||||| |||| || |||| |||  |||| | ||  |||||||||| ||
Db      21652 ATTTTATTTTATTTTTTGAGACAGAGTCTGGCTGTGTCACCCAGGCTGGAGTGCAGTGGC 21711

Qy        139 GTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCT 198
              |||   |||  ||| |  ||     ||||| ||||||| ||| || ||||||||||
Db      21712 ACGATCTCGGCTCGCTGCAACCTCTGCCTCCTGGGTTCAAGTGATTCTCCTGCCTCAGCC 21771

Qy        199 TCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCTGATTTTTAAATTTTTTT 258
              |||||||||||| || || | |  || | |||| || ||||| ||||||  ||||||
Db      21772 TCCCAGGTAGCTGGGATTACAAGCACT-CCACCACGCCCAGCTAATTTTTGTATTTTTAG 21830

Qy        259 TGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATC 310
              || |||   ||         |||| |||||||||| | ||||||||||||  | ||||||
Db      21831 TGGAGATGGGGTTTCACTATGTTGGCCAGGCTGGTTTGGAACTCCTGGCCTCACGTGATC 21890

Qy        311 CTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
              | ||||| ||| ||||||||    ||  || ||||| ||    |||| | | ||
Db      21891 CACCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGCACCTG 21945


RESULT 10
ADQ64936
ID   ADQ64936 standard; cDNA; 3019 BP.
XX
AC   ADQ64936;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Novel human cDNA sequence #2097.
XX
KW   ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
KW   cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW   neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   EP1440981-A2.
XX
PD   28-JUL-2004.
XX
PF   21-JAN-2004; 2004EP-00001196.
XX
PR   21-JAN-2003; 2003JP-00102206.
PR   09-MAY-2003; 2003JP-00131392.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI   Yamamoto J, Isono Y, Nagai K, Irie R;
XX
DR   WPI; 2004-535376/52.
DR   P-PSDB; ADQ67124.
XX
PT   Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 2097; 2449pp; English.
XX
CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotides and polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a nucleotide
CC   sequence of the invention.
XX
SQ   Sequence 3019 BP; 833 A; 670 C; 785 G; 731 T; 0 U; 0 Other;

  Query Match             39.8%;  Score 145.2;  DB 12;  Length 3019;
  Best Local Similarity   71.8%;  Pred. No. 4.7e-29;
  Matches  221;  Conservative    0;  Mismatches   78;  Indels    9;  Gaps    2;

Qy         64 CTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGA 123
              ||| |  | | |||  |  |  ||| ||||| |||| |||||||  ||| || ||| |
Db       1066 CTTGTTTTCTTTTTCTTTCTTTTTTTTTTTTTGGAGACAGGGTCTCACTCTGTTATCCAAG 1125

Qy        124 TTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGAT 183
              |||||||||| ||  | |||| |||| |||| || ||||| |||||| ||||||| |||
Db       1126 CTGGAGTGCAGTGGCATGATCACAGCTCACTGCAGCCTTGACTTCCTGGGTTCAAGTGAT 1185

Qy        184 CCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATG-TGCTGCTACCATGCACAGCTG 242
              ||| |  ||||||  |||   ||||||| ||||| | |     | | ||||||| |||||
Db       1186 CCTCCCACCTCAGTCTCCTGAGTAGCTGGGACTACAGGCATGTACCACCATGCTCAGCTA 1245

Qy        243 ATTTTTAAATTTTTTTTGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTC 294
              || |||||||||||| |  ||| | ||         ||||||||||||||||||||||||
Db       1246 ATCTTTAAATTTTTTGTAGAGACAGGGTCTCACTTTGTTGCCCAGGCTGGTCTTGAACTC 1305

Qy        295 CTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTC 354
              |||   | | ||||| |||||||  ||||||||||||     ||  || ||||| ||    |
Db       1306 CTGAGTTCAAGTGATTCTCCTGCCTTGACCTCCCAAAGTGCTGGGATTACAGGTGTGAGC 1365

Qy        355 CACCACGC 362
              |||||| |
Db       1366 CACCACAC 1373


RESULT 11
ABZ58995
ID   ABZ58995 standard; DNA; 10735 BP.
XX
AC   ABZ58995;
XX
DT   28-APR-2003  (first entry)
XX
DE   Human oncosuppressive gene (DRAGO) fragment.
XX
KW   Oncosuppressive; apoptotic; p53; p73; cytostatic; gene therapy; tumour;
KW   DRAGO; human; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003006498-A2.
XX
PD   23-JAN-2003.
XX
PF   09-JUL-2002; 2002WO-EP007625.
XX
```

```
PR   10-JUL-2001; 2001IT-MI001465.
XX
PA   (NOVU-) NOVUSPHARMA SPA.
XX
PI   Broggini M,  D'incalci M;
XX
DR   WPI; 2003-221715/21.
XX
PT   New oncosuppressive polypeptide, useful for preparing a medicament for
PT   treating tumors.
XX
PS   Claim 3; Page 34-37; 42pp; English.
XX
CC   The invention relates to oncosuppressive polypeptides and encoding
CC   polynucleotides. The oncosuppressive gene is involved in apoptotic
CC   process and is regulated by p53 and p73. The oncosuppressive
CC   polynucleotides are useful for preparing a medicament for treating
CC   tumour. The present sequence represents a human oncosuppressive gene
CC   (DRAGO) fragment, located upstream of the first exon
XX
SQ   Sequence 10735 BP; 2720 A; 2485 C; 2630 G; 2900 T; 0 U; 0 Other;

  Query Match            39.7%;  Score 145;  DB 8;  Length 10735;
  Best Local Similarity   67.8%;  Pred. No. 7.7e-29;
  Matches  234;  Conservative    0;  Mismatches  105;  Indels    6;  Gaps    2;

Qy         19 TTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTA 78
              |||||| |      |    || | | |||    | | || |||| | |||| | || || |
Db        329 TTACTTTAACAACTATAATTTAATTTTACATATATATAATTTAATTTTTTTAATTTTTAA 388

Qy         79 GTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGT 138
              | ||    | ||||| |||| || ||||||||| |||| |||   ||||||| || ||
Db        389 TTTTTTTGGTTTTTTTTGAGACAGAGTCTTGCTCTGTCACCTAAGCTGGAGTGTAGTGGC 448

Qy        139 GTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCT 198
                ||||| |||||||| ||||| |||  ||| ||| ||||   ||||| |  |||||||
Db        449 ACAATCACAGCTTACTTTAGTCCTGACCTCCCGAGCTCAAATAATCCTCCCACCTCAGCC 508

Qy        199 TCCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGATTTTTAAATTTTTT 257
              |||   ||||||| ||| | | | |  ||| |||||||   ||| ||||||  |||||||
Db        509 TCCTGAGTAGCTGGGACCACAGGCGTGTGCCACCATGCCTGGCTAATTTTTGTATTTTTT 568

Qy        258 TTGTAGAG-----ATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCT 312
              |  ||||     |    ||||||||||||||||  |||| |||  || | |||||||
Db        569 GTAAAGAGGTCTCACTCTGTTGCCCAGGCTGGTCTCAAACTTCTGAGCTCAAGTGATCCT 628

Qy        313 CCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCAC 357
              ||||| ||| ||||||||    ||  || |||||  |    ||||
Db        629 CCTGCTTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCAC 673


RESULT 12
AAK88992/c
ID   AAK88992 standard; DNA; 11216 BP.
XX
AC   AAK88992;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 2568.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001324.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
```

PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
PS Disclosure; SEQ ID NO 2568; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 11216 BP; 3857 A; 1984 C; 1812 G; 3563 T; 0 U; 0 Other;

Query Match 39.3%; Score 143.6; DB 4; Length 11216;
Best Local Similarity 71.6%; Pred. No. 1.9e-28;
Matches 202; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

```
Qy        82 TTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTA 141
             ||  ||| |||||||||| | |||| ||||| ||||   ||  | ||||||| |||| |
Db      8240 TTTGTTTGTTTTTAGAGATAAGGTCCTGCTCTGTCACTCAGGCTAGAGTGCAGTGGTGCA 8181

Qy       142 ATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCC 201
             ||||| ||| |||| || |||||| | ||| | |||||  ||||| |||||||||| ||||
Db      8180 ATCATTGCTCACTGCAGCCTTGAACTTCTGGGCTCAAGCAATCCTCCTGCCTCAGATTCC 8121

Qy       202 CAGGTAGCTGAGACTATATGTGC-TGCTACCATGCACAGCTGATTTTTAAATTTTTTTG 260
             |||||||||| ||||||| |  | ||| ||||| | ||||| ||||||           |
Db      8120 CAGGTAGCTGGGACTATAGGCACATGCCACCATCCCCAGCTAATTTTTTGTAGAGACACG 8061

Qy       261 TAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTT 320
               | | ||||||||||||||||| ||||| |||| || | |   ||||||||| ||
Db      8060 GTCTCACTGTGTTGCCCAGGCTGGTCTCGAACTTCTGGGCTCAAGCTGTCCTCCTGCCTT 8001

Qy       321 GACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGC 362
             |||| ||||    ||  || ||||||      ||||||| |
Db      8000 AACCTTCCAAAATACTGGGATTACAGACATGAGCCACCACAC 7959
```

RESULT 13
AAS39593/c
ID AAS39593 standard; DNA; 11216 BP.
XX
AC AAS39593;
XX
DT 17-DEC-2001 (first entry)
XX
DE Genomic sequence #12 encoding human colon associated polypeptide.
XX
KW Human; colon cancer; congenital abnormality; infection; colitis;
KW inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW sigmoid disease; antibacterial; antiviral; antiinflammatory; cytostatic;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200155302-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001240.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465567/50.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the colon including colon cancers and also for
PT testing and detection e.g. diagnosis.
XX
PS Disclosure; SEQ ID NO 490; 562pp; English.
XX
CC The present invention relates to the isolation of novel human colon
CC associated polypeptides (AAU22468-AAU22701), and the cDNA and genomic
CC sequences encoding for them. The sequences of the invention are useful in
CC the diagnosis, treatment, prevention and/or prognosis of disorders of the
CC colon including colon cancer, congenital abnormalities (e.g. atresia and
CC stenosis), bacterial and viral infections, inflammatory bowel disease
CC (IBD), neoplastic cell disorders (e.g. polyps and adenomas, intestinal
CC inflammatory disorders, colitis, colonic inflammation, diarrhoea and

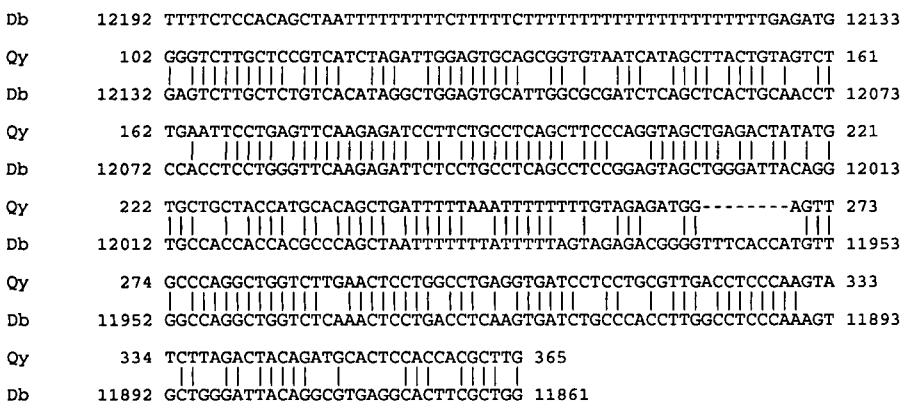
```
CC   dysentery, malabsorption syndromes (e.g. lactose intolerance), intestinal
CC   obstruction and sigmoid diseases. The polynucleotides sequences of the
CC   invention can also be used in gene therapy. AAS39582-AAS40060 represent
CC   DNA sequences encoding for the novel human colon associated polypeptides
CC   of the invention. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 11216 BP; 3857 A; 1984 C; 1812 G; 3563 T; 0 U; 0 Other;

  Query Match             39.3%;  Score 143.6;  DB 5;  Length 11216;
  Best Local Similarity   71.6%;  Pred. No. 1.9e-28;
  Matches  202;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy         82 TTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTA 141
              ||  ||| ||||||||||  | |||| ||||| ||||   ||  | ||||||| |||| |
Db       8240 TTTGTTTGTTTTTAGAGATAAGGTCCTGCTCTGTCACTCAGGCTAGAGTGCAGTGGTGCA 8181

Qy        142 ATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCC 201
              ||||| ||| |||| || |||||| | ||| | |||||  ||||| |||||||||| ||||
Db       8180 ATCATTGCTCACTGCAGCCTTGAACTTCTGGGCTCAAGCAATCCTCCTGCCTCAGATTCC 8121

Qy        202 CAGGTAGCTGAGACTATATGTGC-TGCTACCATGCACAGCTGATTTTTAAATTTTTTTTG 260
              |||||||||| ||||||| |   | ||| ||||| | ||||| ||||||           |
Db       8120 CAGGTAGCTGGGACTATAGGCACATGCCACCATCCCCAGCTAATTTTTTTGTAGAGACACG 8061

Qy        261 TAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTT 320
                |  | ||||||||||||||||| ||||| |||| || | |   |||||||| ||
Db       8060 GTCTCACTGTGTTGCCCAGGCTGGTCTCGAACTTCTGGGCTCAAGCTGTCCTCCTGCCTT 8001

Qy        321 GACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGC 362
               |||| ||||    ||  || ||||||      ||||||| |
Db       8000 AACCTTCCAAAATACTGGGATTACAGACATGAGCCACCACAC 7959


RESULT 14
ADB32553/c
ID   ADB32553 standard; DNA; 11216 BP.
XX
AC   ADB32553;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human novel colon related polypeptide DNA SEQ ID NO 490.
XX
KW   gene therapy; cancer; liver disorder; hepatitis; neural disorder;
KW   Alzheimer's disease; human; colon; ds.
XX
OS   Homo sapiens.
XX
PN   US2003050231-A1.
XX
PD   13-MAR-2003.
XX
PF   17-JAN-2001; 2001US-00764872.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
```

```
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-625420/59.
XX
PT   New nucleic acid molecule, useful for preparing a medicament for
PT   preventing, treating or ameliorating a medical condition e.g. cancer,
PT   liver disorders or neural disorders.
XX
PS   Disclosure; SEQ ID NO 490; 216pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule encoding a
CC   polypeptide. The nucleic acid is useful for preparing a medicament for
CC   preventing, treating or ameliorating a medical condition e.g. cancer,
CC   liver disorders such as hepatitis or neural disorders such as Alzheimer's
CC   disease. The present sequence represents a human novel colon related
CC   polypeptide DNA. Note: The sequence data for this patent did not form
CC   part of the printed specification but was obtained in electronic format
CC   directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030050231.
XX
SQ   Sequence 11216 BP; 3857 A; 1984 C; 1812 G; 3563 T; 0 U; 0 Other;

  Query Match           39.3%;  Score 143.6;  DB 9;  Length 11216;
  Best Local Similarity   71.6%;  Pred. No. 1.9e-28;
  Matches  202;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy         82 TTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTA 141
              ||  ||| ||||||||||| | |||| ||||| ||||    ||   | ||||||| |||| |
Db       8240 TTTGTTTGTTTTTAGAGATAAGGTCCTGCTCTGTCACTCAGGCTAGAGTGCAGTGGTGCA 8181

Qy        142 ATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCC 201
              ||||| ||| |||| || |||||| | ||| | |||||  ||||| ||||||||| ||||
Db       8180 ATCATTGCTCACTGCAGCCTTGAACTTCTGGGCTCAAGCAATCCTCCTGCCTCAGATTCC 8121

Qy        202 CAGGTAGCTGAGACTATATGTGC-TGCTACCATGCACAGCTGATTTTTAAATTTTTTTTG 260
              |||||||||| ||||||| |  | ||| ||||| | ||||| ||||||           |
Db       8120 CAGGTAGCTGGGACTATAGGCACATGCCACCATCCCCAGCTAATTTTTTGTAGAGACACG 8061

Qy        261 TAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTT 320
                | | ||||||||||||||||| ||||| |||| || | |   ||||||||| ||
Db       8060 GTCTCACTGTGTTGCCCAGGCTGGTCTCGAACTTCTGGGCTCAAGCTGTCCTCCTGCCTT 8001

Qy        321 GACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGC 362
               |||| ||||    ||  || ||||||       ||||||| |
Db       8000 AACCTTCCAAAATACTGGGATTACAGACATGAGCCACCACAC 7959


RESULT 15
AAK72853/c
ID   AAK72853 standard; DNA; 14540 BP.
XX
AC   AAK72853;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27665.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
```

PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27665; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 14540 BP; 4313 A; 3425 C; 3025 G; 3777 T; 0 U; 0 Other;

Query Match 39.1%; Score 142.8; DB 4; Length 14540;
Best Local Similarity 66.9%; Pred. No. 3.3e-28;
Matches 222; Conservative 0; Mismatches 102; Indels 8; Gaps 1;

Qy 42 TCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAA 101



Search completed: August 30, 2005, 04:04:03
Job time : 374 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 03:39:08 ; Search time 136 Seconds
(without alignments)
4391.478 Million cell updates/sec

Title: US-08-731-499-9_COPY_10001_10365
Perfect score: 365
Sequence: 1 TTTGTGGTCCTCCAAGGCTT.....GATGCACTCCACCACGCTTG 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	365	100.0	10365	4	US-08-892-695-9	Sequence 9, Appli
	2	365	100.0	20022	4	US-09-949-016-12604	Sequence 12604, A
	3	365	100.0	20023	4	US-09-949-016-16004	Sequence 16004, A
	4	149.2	40.9	38538	4	US-09-949-016-13150	Sequence 13150, A
	5	146.8	40.2	197131	4	US-09-949-016-12675	Sequence 12675, A
	6	146.8	40.2	197132	4	US-09-949-016-17170	Sequence 17170, A
c	7	145.2	39.8	22010	4	US-09-949-016-15960	Sequence 15960, A
	8	144.8	39.7	601	4	US-09-949-016-51283	Sequence 51283, A
	9	144.8	39.7	52971	4	US-09-949-016-16452	Sequence 16452, A
	10	144.8	39.7	73295	4	US-09-949-016-15151	Sequence 15151, A
c	11	144.8	39.7	121427	4	US-09-949-016-11950	Sequence 11950, A
c	12	144.8	39.7	121433	4	US-09-949-016-13230	Sequence 13230, A
c	13	144.4	39.6	25736	4	US-09-949-016-15090	Sequence 15090, A
c	14	144.4	39.6	25755	4	US-09-949-016-12351	Sequence 12351, A
c	15	144.2	39.5	601	4	US-09-949-016-121870	Sequence 121870,
c	16	143	39.2	33392	4	US-09-949-016-15172	Sequence 15172, A
	17	142.4	39.0	24665	4	US-09-949-016-17134	Sequence 17134, A
	18	142.2	39.0	36759	4	US-09-949-016-12216	Sequence 12216, A
	19	142.2	39.0	36760	4	US-09-949-016-14021	Sequence 14021, A
c	20	142	38.9	601	4	US-09-949-016-157469	Sequence 157469,
c	21	142	38.9	601	4	US-09-949-016-157576	Sequence 157576,
	22	142	38.9	87734	4	US-09-949-016-17521	Sequence 17521, A
c	23	141	38.6	126254	4	US-09-949-016-15341	Sequence 15341, A
	24	141	38.6	235452	4	US-09-949-016-13675	Sequence 13675, A
c	25	140.8	38.6	601	4	US-09-949-016-141029	Sequence 141029,
c	26	140.8	38.6	601	4	US-09-949-016-141030	Sequence 141030,
c	27	140.8	38.6	601	4	US-09-949-016-157470	Sequence 157470,
c	28	140.8	38.6	601	4	US-09-949-016-157577	Sequence 157577,
c	29	140.8	38.6	126176	4	US-09-949-016-16137	Sequence 16137, A
c	30	140.8	38.6	126176	4	US-09-949-016-16138	Sequence 16138, A
	31	140.4	38.5	601	4	US-09-949-016-121481	Sequence 121481,
c	32	140.4	38.5	601	4	US-09-949-016-157471	Sequence 157471,
c	33	140.4	38.5	601	4	US-09-949-016-157578	Sequence 157578,
	34	140.4	38.5	19253	4	US-09-949-016-15131	Sequence 15131, A
	35	140.4	38.5	83178	4	US-09-949-016-14606	Sequence 14606, A
	36	140.2	38.4	601	4	US-09-949-016-89655	Sequence 89655, A
	37	140.2	38.4	601	4	US-09-949-016-89666	Sequence 89666, A
	38	140.2	38.4	601	4	US-09-949-016-89677	Sequence 89677, A
	39	140.2	38.4	601	4	US-09-949-016-89688	Sequence 89688, A
c	40	140.2	38.4	152582	4	US-09-949-016-12086	Sequence 12086, A
c	41	140.2	38.4	152583	4	US-09-949-016-17390	Sequence 17390, A
c	42	140.2	38.4	152583	4	US-09-949-016-17391	Sequence 17391, A
c	43	140	38.4	20303	1	US-08-370-975B-6	Sequence 6, Appli
c	44	140	38.4	26764	1	US-08-370-975B-1	Sequence 1, Appli
	45	139.8	38.3	601	4	US-09-949-016-89656	Sequence 89656, A

ALIGNMENTS

RESULT 1
US-08-892-695-9
; Sequence 9, Application US/08892695A
; Patent No. 6808878
; GENERAL INFORMATION:
; APPLICANT: Gray, Joe W
; APPLICANT: Collins, Collin
; APPLICANT: Hwang, Soo In
; APPLICANT: Godfrey, Tony
; APPLICANT: Kowel, David
; APPLICANT: Rommens, Johanna
; TITLE OF INVENTION: GENES FROM THE 20Q13 AMPLICON AND THEIR USES
; FILE REFERENCE: 2500.124US3
; CURRENT APPLICATION NUMBER: US/08/892,695A
; CURRENT FILING DATE: 1997-07-15
; EARLIER APPLICATION NUMBER: 08/785,532
; EARLIER FILING DATE: 1997-01-17
; EARLIER APPLICATION NUMBER: 08/731,499
; EARLIER FILING DATE: 1996-10-16
; EARLIER APPLICATION NUMBER: 08/680,395
; EARLIER FILING DATE: 1996-07-15
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 10365
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Genomic
; OTHER INFORMATION: Sequence encoding ZABC1
; Patent No. 6808878
US-08-892-695-9

Query Match 100.0%; Score 365; DB 4; Length 10365;
Best Local Similarity 100.0%; Pred. No. 1.1e-100;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10001 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 10060

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10061 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 10120

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 10180

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 10240

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 10300

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 10360

Qy        361 GCTTG 365
              |||||
Db      10361 GCTTG 10365
```

RESULT 2
US-09-949-016-12604
; Sequence 12604, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12604
; LENGTH: 20022
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12604

Query Match 100.0%; Score 365; DB 4; Length 20022;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9531 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 9590

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9591 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 9650

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9651 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 9710

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9711 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 9770

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9771 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 9830

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9831 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 9890

Qy        361 GCTTG 365
              |||||
Db       9891 GCTTG 9895
```

RESULT 3
US-09-949-016-16004
; Sequence 16004, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16004
; LENGTH: 20023
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16004

Query Match 100.0%; Score 365; DB 4; Length 20023;
Best Local Similarity 100.0%; Pred. No. 1.5e-100;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9531 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 9590

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9591 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 9650

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9651 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 9710

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9711 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 9770

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9771 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 9830

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9831 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 9890

Qy        361 GCTTG 365
              |||||
Db       9891 GCTTG 9895
```

RESULT 4
US-09-949-016-13150
; Sequence 13150, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768

```
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13150
;   LENGTH: 38538
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(38538)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13150

  Query Match            40.9%;  Score 149.2;  DB 4;  Length 38538;
  Best Local Similarity  75.4%;  Pred. No. 8.1e-35;
  Matches  214;  Conservative    0;  Mismatches   63;  Indels    7;  Gaps    2;

Qy         83 TACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAA 142
              || ||  |||| ||||| |||||||  ||| |||||  ||| || |||||||  || || |
Db       5869 TAATTGTTTTTAAGAGACAGGGTCTCACTCTGTCATTCAGACTGAAGTGCAGTGGAGTGA 5928

Qy        143 TCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCC 202
              |||||||| |||| || || ||| ||||| | ||||| |||||| |||||||||| || |
Db       5929 TCATAGCTCACTGCAGCCTAGAACTCCTGGGCTCAAGTGATCCTCCTGCCTCAGCCTCTC 5988

Qy        203 AGGTAGCTGAGACTATATGTGC-TGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGT 261
              |||||||||||||||| |  | ||| ||||||| ||||| ||||| ||| |||||||||
Db       5989 CAGTAGCTGAGACTATAGGCACATGCCACCATGCCCAGCTAATTTTAAAACTTTTTTTGT 6048

Qy        262 AGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTG 321
              |  ||      |||  |||||||||||||||||||||| | | |  ||||||| || ||
Db       6049 AAGGA------TGCTGAGGCTGGTCTTGAACTCCTGGCATCAAGCAATCCTCCCGCCTTA 6102

Qy        322 ACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
              |||||||| |  ||  || || ||  |    |||||| |||||
Db       6103 GCCTCCCAAATTCCTGGGATTAGAGGCGTGAGCCACCATGCTTG 6146


RESULT 5
US-09-949-016-12675
; Sequence 12675, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12675
;   LENGTH: 197131
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(197131)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12675

  Query Match            40.2%;  Score 146.8;  DB 4;  Length 197131;
  Best Local Similarity  69.6%;  Pred. No. 9.3e-34;
  Matches  231;  Conservative    0;  Mismatches   92;  Indels    9;  Gaps    2;

Qy         40 ACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGA 99
              |||| |      | ||| | |   | ||||  ||  ||| || ||  ||| ||||| ||||
Db     134568 ACTCATTGTATTCTGTGTGTGTGTTTTTTTTGTTTTTGTTTTTGTTTTTTTTTTGAGA 134627

Qy        100 AAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGT 159
              || ||||  ||| |||| | ||  |||||||||| ||||| |||  |||| |||| |
Db     134628 CAGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAAC 134687

Qy        160 CTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATA 219
              ||  |  ||||| |||||||  || || |||||| ||| ||||| |||||||| |||||||
Db     134688 CTCCACCTCCTGGGTTCAAGCAATTCTCCTGCCTTAGCCTCCCAAGTAGCTGGGACTATA 134747

Qy        220 TGTGC-TGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG--------A 270
              | ||  || |||| || ||||| ||||||  ||||||  |  ||| | ||
Db     134748 GGCGCACGCCACCACGCCCAGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTTACCAC 134807

Qy        271 GTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAA 330
              |||| |||| | |||||||||||||||| ||| ||||||||||||  | | | ||||||||
Db     134808 GTTGGCCAGACCGGTCTTGAACTCCTGACCTCAGGTGATCCTCCCACCTCGGCCTCCCAA 134867

Qy        331 GTATCTTAGACTACAGATGCACTCCACCACGC 362
              |   ||  || ||||| ||    |||||  ||
Db     134868 GGTGCTGGGATTACAGGTGTGAGCCACCGTGC 134899


RESULT 6
US-09-949-016-17170
; Sequence 17170, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17170
;   LENGTH: 197132
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(197132)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17170

  Query Match            40.2%;  Score 146.8;  DB 4;  Length 197132;
  Best Local Similarity  69.6%;  Pred. No. 9.3e-34;
  Matches  231;  Conservative    0;  Mismatches   92;  Indels    9;  Gaps    2;

Qy         40 ACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGA 99
              |||| |      | ||| | |   | ||||  ||  ||| || ||  ||| ||||| ||||
Db     134568 ACTCATTGTATTCTGTGTGTGTGTTTTTTTTGTTTTTGTTTTTGTTTTTTTTTTGAGA 134627

Qy        100 AAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGT 159
              || ||||  ||| |||| | ||  |||||||||| ||||| |||  |||| |||| |
Db     134628 CAGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAAC 134687

Qy        160 CTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATA 219
              ||  |  ||||| |||||||  || || |||||| ||| ||||| |||||||| |||||||
Db     134688 CTCCACCTCCTGGGTTCAAGCAATTCTCCTGCCTTAGCCTCCCAAGTAGCTGGGACTATA 134747

Qy        220 TGTGC-TGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG--------A 270
              | ||  || |||| || ||||| ||||||  ||||||  |  ||| | ||
Db     134748 GGCGCACGCCACCACGCCCAGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTTACCAC 134807
```

```
Qy        271 GTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAA 330
              |||| |||| | ||||||||||||||| ||| ||||||||||||  | | | ||||||||
Db     134808 GTTGGCCAGACCGGTCTTGAACTCCTGACCTCAGGTGATCCTCCCACCTCGGCCTCCCAA 134867

Qy        331 GTATCTTAGACTACAGATGCACTCCACCACGC 362
              |   ||  || ||||| ||    |||||  ||
Db     134868 GGTGCTGGGATTACAGGTGTGAGCCACCGTGC 134899


RESULT 7
US-09-949-016-15960/c
; Sequence 15960, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15960
;   LENGTH: 22010
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15960

  Query Match             39.8%;  Score 145.2;  DB 4;  Length 22010;
  Best Local Similarity   67.7%;  Pred. No. 1e-33;
  Matches  222;  Conservative    0;  Mismatches   98;  Indels    8;  Gaps    1;

Qy         43 CTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAG 102
              | |||   ||| | |||| || | ||   || | |  | ||   || | ||| |||| ||
Db      18210 CATAATTTTGTCTTTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGAGACAG 18151

Qy        103 GGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTT 162
              ||| |  ||| ||  || ||  |||||||||| |||  ||| || ||| |||| || |||
Db      18150 GGTGTCACTCTGTTGTCCAGGCTGGAGTGCAGTGGTACAATAATGGCTCACTGCAGCCTT 18091

Qy        163 GAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGT 222
               |  ||||| | |||||  ||||| |||||||||| |||| ||||||||| || | | ||
Db      18090 AACCTCCTGGGCTCAAGCAATCCTCCTGCCTCAGCCTCCCTGGTAGCTGAAACCACAGGT 18031

Qy        223 GCTGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG--------AGTTG 274
              ||   |  |      | ||| |  ||  ||| ||| |  ||| | ||          ||||
Db      18030 GCATGTCACCACACCGGCTATTGCTTGTATTGTTTGTAGAGACAAGGTCTCACTATGTTG 17971

Qy        275 CCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTAT 334
              |||||||||||||||||||||||  |   |||||||||||| |   |||||| ||||
Db      17970 CCCAGGCTGGTCTTGAACTCCTGGGTTCCAGTGATCCTCCTGCCTCAGCCTCCCTAGTAG 17911

Qy        335 CTTAGACTACAGATGCACTCCACCACGC 362
              || | || |||| ||||   |||||| |
Db      17910 CTGAAACCACAGGTGCATGTCACCACAC 17883


RESULT 8
US-09-949-016-51283
; Sequence 51283, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 51283
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-51283

  Query Match             39.7%;  Score 144.8;  DB 4;  Length 601;
  Best Local Similarity   67.2%;  Pred. No. 2.5e-34;
  Matches  238;  Conservative    0;  Mismatches  107;  Indels    9;  Gaps    2;

Qy         20 TACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAG 79
              |||| ||  ||      ||    |  || |  || | |   |    |  ||||| |||
Db        226 TACTAAATACTACCTTTTCTTATGCAAAACTAGTATTTCAAAAATATCAGATTTCTTTTC 285

Qy         80 TCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTG 139
              | ||  ||| ||||| |||| || ||||  ||| |||||| ||  |||||||||| || |
Db        286 TTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTCATCCAGGCTGGAGTGCAGTGGCG 345

Qy        140 TAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTT 199
              | ||| | ||| |||| |  |||    ||||| ||||||| ||| || ||||||||| |
Db        346 TGATCTTGGCTCACTGCAACCTTTGCCTCCTGGGTTCAAGTGATTCTCCTGCCTCAGCCT 405

Qy        200 CCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGATTTTTAAATTTTTTT 258
              |||  |||||||||| || | ||| |  | ||||||| | ||| ||||||  ||||||
Db        406 CCCGAGTAGCTGAGATTACAGGTGCCCACCACCATGCCCGGCTAATTTTTGTATTTTTAG 465

Qy        259 TGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATC 310
              |  |||   ||         |||| ||||||||||||||||||||| ||| | ||||||
Db        466 TAAAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAAGTGATC 525

Qy        311 CTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTT 364
              | || || | | ||||||||    ||  || |||||     ||||||||||
Db        526 CGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAACCACCACGCTT 579


RESULT 9
US-09-949-016-16452
; Sequence 16452, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16452
;   LENGTH: 52971
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(52971)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16452
```

```
  Query Match             39.7%;  Score 144.8;  DB 4;  Length 52971;
  Best Local Similarity   70.2%;  Pred. No. 2e-33;
  Matches  212;  Conservative    0;  Mismatches   82;  Indels    8;  Gaps    1;

Qy         72 TTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTG 131
              || |||  | ||  ||| ||||| |||  | |||| |||  ||   ||||  |||||||
Db      21346 TTTTTTTTTTTTTTTTTTTTTTTTGAGTTGGAGTCTCGCTGTGTTGCCTAGGCTGGAGTG 21405

Qy        132 CAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGC 191
              |||| | | |||| ||||| |||| || ||  |  ||||| ||||||   || || ||||
Db      21406 CAGCAGGGCAATCTTAGCTCACTGCAGCCTCCACCTCCTGGGTTCAAACAATTCTCCTGC 21465

Qy        192 CTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCTGATTTTTAAA 251
              |||||| ||||| ||||||| ||||| | |    ||| ||||| | ||||| |||||||
Db      21466 CTCAGCCTCCCAAGTAGCTGGGACTACAGGCTGTGCCACCATACCCAGCTAATTTTTATG 21525

Qy        252 TTTTTTTTGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGA 303
              |||||  |  ||| | ||          |||| ||||||||||||||||||||| ||| |
Db      21526 TTTTTAGTAGAGACAGGGTTTTGCCATGTTGTCCAGGCTGGTCTTGAACTCCTGACCTCA 21585

Qy        304 GGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCT 363
              ||||||| ||||  |  |||||||||     ||  || |||||| |    ||||||  |
Db      21586 AGTGATCCACCTGTCTCCACCTCCCAAAGTGCTGGGATTACAGACGTGGGCCACCATCCC 21645

Qy        364 TG 365
              ||
Db      21646 TG 21647


RESULT 10
US-09-949-016-15151
; Sequence 15151, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15151
;   LENGTH: 73295
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(73295)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15151

  Query Match             39.7%;  Score 144.8;  DB 4;  Length 73295;
  Best Local Similarity   72.8%;  Pred. No. 2.4e-33;
  Matches  217;  Conservative    0;  Mismatches   72;  Indels    9;  Gaps    2;

Qy         75 TTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAG 134
              ||   | ||  ||| ||||| |||| |||||||| ||| |||| | ||| ||||||||||
Db      49031 TTGTTTTTTTTTTTTTTTTTTTGAGACAGGGTCTTACTCTGTCACCCAGACTGGAGTGCAG 49090

Qy        135 CGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTC 194
              ||   |||||||||| ||||||| |||||   ||   | |||||| || || |  || |
Db      49091 TGGCACAATCATAGCTCACTGTAGCCTTGACCGCCCAGGCTCAAGACATTCTCCCACCCC 49150

Qy        195 AGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGATTTTTAAATT 253
              ||| ||||| ||||||| ||||| | |||  ||| ||||||| ||||| ||||||  |||
Db      49151 AGCCTCCCAAGTAGCTGGGACTACAGGTGTGTGCCACCATGCCCAGCTAATTTTTGTATT 49210

Qy        254 TTTTTTGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGG 305
              |||| |  | | | ||          |||||||||||||||||||||||||||| |  |||
Db      49211 TTTTGTAGAAACAAGGTTTCCCCATGTTGCCCAGGCTGGTCTTGAACTCCTGGGCACAGG 49270

Qy        306 TGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCT 363
              ||||  ||||| | | |||||||||     ||  || ||||||||    ||||||| ||
Db      49271 GGATCTGCCTGCCTGGGCCTCCCAAAGTGCTGGGATTACAGATGGGAGCCACCACACT 49328


RESULT 11
US-09-949-016-11950/c
; Sequence 11950, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11950
;   LENGTH: 121427
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-11950

  Query Match             39.7%;  Score 144.8;  DB 4;  Length 121427;
  Best Local Similarity   67.2%;  Pred. No. 3e-33;
  Matches  238;  Conservative    0;  Mismatches  107;  Indels    9;  Gaps    2;

Qy         20 TACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAG 79
              |||| ||  ||      ||     |  || |  || | |   |     |  ||||| |||
Db      33363 TACTAAATACTACCTTTTCTTATGCAAAACTAGTATTTCAAAAATATCAGATTTCTTTTC 33304

Qy         80 TCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTG 139
              | ||  ||| ||||| |||| || ||||  ||| ||||||  ||  |||||||||  || |
Db      33303 TTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTCATCCAGGCTGGAGTGCAGTGGCG 33244

Qy        140 TAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTT 199
              | ||| | ||| |||| |  |||    ||||| ||||||| ||| || ||||||||| |
Db      33243 TGATCTTGGCTCACTGCAACCTTTGCCTCCTGGGTTCAAGTGATTCTCCTGCCTCAGCCT 33184

Qy        200 CCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGATTTTTAAATTTTTTT 258
              |||  |||||||||| || | ||| |  | ||||||| | ||| ||||||   ||||||
Db      33183 CCCGAGTAGCTGAGATTACAGGTGCCCACCACCATGCCCGGCTAATTTTTGTATTTTTAG 33124

Qy        259 TGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATC 310
              |  |||   ||         |||| ||||||||||||||||||||| ||| | ||||||
Db      33123 TAAAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAAGTGATC 33064

Qy        311 CTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTT 364
              | || || | | ||||||||     ||  || |||||        ||||||||||
Db      33063 CGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAACCACCACGCTT 33010


RESULT 12
US-09-949-016-13230/c
; Sequence 13230, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
```

```
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13230
;   LENGTH: 121433
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-13230

  Query Match             39.7%;  Score 144.8;  DB 4;  Length 121433;
  Best Local Similarity   67.2%;  Pred. No. 3e-33;
  Matches  238;  Conservative    0;  Mismatches  107;  Indels    9;  Gaps    2;

Qy         20 TACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAG 79
              |||| ||  ||      ||     |   || |  || | |   |     |  ||||| |||
Db      33362 TACTAAATACTACCTTTTCTTATGCAAAACTAGTATTTCAAAAATATCAGATTTCTTTTC 33303

Qy         80 TCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTG 139
              | ||  ||| ||||| |||| || ||||  ||| |||||| ||  |||||||||| || |
Db      33302 TTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTCATCCAGGCTGGAGTGCAGTGGCG 33243

Qy        140 TAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTT 199
              | ||| | ||| |||| |  |||     ||||| ||||||| ||| || ||||||||||  |
Db      33242 TGATCTTGGCTCACTGCAACCTTTGCCTCCTGGGTTCAAGTGATTCTCCTGCCTCAGCCT 33183

Qy        200 CCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGATTTTTAAATTTTTTT 258
              |||  |||||||||| || | ||| |  | ||||||| | ||| ||||||  ||||||
Db      33182 CCCGAGTAGCTGAGATTACAGGTGCCCACCACCATGCCCGGCTAATTTTTGTATTTTTAG 33123

Qy        259 TGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATC 310
              |  |||   ||         |||| ||||||||||||||||||||| ||| | ||||||
Db      33122 TAAAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCTCAAGTGATC 33063

Qy        311 CTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTT 364
              | || || | | |||||||||    ||  || |||||       ||||||||||
Db      33062 CGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAACCACCACGCTT 33009


RESULT 13
US-09-949-016-15090/c
; Sequence 15090, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15090
;   LENGTH: 25736
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-15090

  Query Match             39.6%;  Score 144.4;  DB 4;  Length 25736;
  Best Local Similarity   70.4%;  Pred. No. 1.9e-33;
  Matches  207;  Conservative    0;  Mismatches   86;  Indels    1;  Gaps    1;

Qy         54 GTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCC 113
              || |  | | | | ||  |  |||  | |   ||| |  || ||||  | ||||  |||
Db      20931 GTCTACTCTACATATGCATATTTTCTTTTCTTTTTTTCCTTTGAGACGGAGTCTCACTCT 20872

Qy        114 GTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAG 173
              |||  | ||  || |||||||  |||| |||  | ||| |||| |  |||   ||||   |
Db      20871 GTCGCCCAGGCTGAAGTGCAGTGGTGCGATCTTGGCTCACTGCAACCTTTGCTTCCCAGG 20812

Qy        174 TTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTACCA 232
              |||||| ||| || |||||||||| ||||  ||||||| || || | | |  ||| ||||
Db      20811 TTCAAGCGATTCTCCTGCCTCAGCCTCCCCAGTAGCTGGGATTACAGGCGTGTGCCACCA 20752

Qy        233 TGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAAC 292
              ||| ||||| ||||||   | ||||| |||||||||  |||| ||||||||||| ||||||
Db      20751 TGCCCAGCTAATTTTTCTGTATTTTTAGTAGAGATGAGGTTGGCCAGGCTGGTTTTGAAC 20692

Qy        293 TCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAG 346
              ||||| ||| | ||||||| ||||| | | ||||||||    ||  || |||||
Db      20691 TCCTGACCTCAAGTGATCCACCTGCCTCGGCCTCCCAAAGCGCTAGGATTACAG 20638


RESULT 14
US-09-949-016-12351/c
; Sequence 12351, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12351
;   LENGTH: 25755
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12351

  Query Match             39.6%;  Score 144.4;  DB 4;  Length 25755;
  Best Local Similarity   70.4%;  Pred. No. 1.9e-33;
  Matches  207;  Conservative    0;  Mismatches   86;  Indels    1;  Gaps    1;

Qy         54 GTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCC 113
              || |  | | | | ||  |  |||  | |   ||| |  || ||||  | ||||  |||
Db      20950 GTCTACTCTACATATGCATATTTTCTTTTCTTTTTTTCCTTTGAGACGGAGTCTCACTCT 20891

Qy        114 GTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAG 173
              |||  | ||  || |||||||  |||| |||  | ||| |||| |  |||   ||||   |
Db      20890 GTCGCCCAGGCTGAAGTGCAGTGGTGCGATCTTGGCTCACTGCAACCTTTGCTTCCCAGG 20831

Qy        174 TTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTACCA 232
              |||||| ||| || |||||||||| ||||  ||||||| || || | | |  ||| ||||
Db      20830 TTCAAGCGATTCTCCTGCCTCAGCCTCCCCAGTAGCTGGGATTACAGGCGTGTGCCACCA 20771

Qy        233 TGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAAC 292
              ||| ||||| ||||||   | ||||| |||||||||  |||| ||||||||||| ||||||
Db      20770 TGCCCAGCTAATTTTTCTGTATTTTTAGTAGAGATGAGGTTGGCCAGGCTGGTTTTGAAC 20711

Qy        293 TCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAG 346
              ||||| ||| | ||||||| ||||| | | ||||||||    ||  || |||||
Db      20710 TCCTGACCTCAAGTGATCCACCTGCCTCGGCCTCCCAAAGCGCTAGGATTACAG 20657
```

```
RESULT 15
US-09-949-016-121870/c
; Sequence 121870, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 121870
;   LENGTH: 601
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-121870

  Query Match             39.5%;  Score 144.2;  DB 4;  Length 601;
  Best Local Similarity   73.5%;  Pred. No. 3.9e-34;
  Matches  214;  Conservative    0;  Mismatches   68;  Indels    9;  Gaps    2;

Qy          82 TTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTA 141
               ||  ||| ||||| |||| |||||||| ||| |||| | ||| |||||||||| ||   |
Db         600 TTTTTTTTTTTTTTGAGACAGGGTCTTACTCTGTCACCCAGACTGGAGTGCAGTGGCACA 541

Qy         142 ATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCC 201
               ||||||||| ||||||| |||||   ||   | |||||| || || |  || |||| |||
Db         540 ATCATAGCTCACTGTAGCCTTGACCGCCCAGGCTCAAGACATTCTCCCACCCCAGCCTCC 481

Qy         202 CAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGATTTTTAAATTTTTTTTG 260
               || ||||||| ||||| | |||  ||| ||||||| ||||| ||||||  ||||||| |
Db         480 CAAGTAGCTGGGACTACAGGTGTGTGCCACCATGCCCAGCTAATTTTTGTATTTTTTGTA 421

Qy         261 TAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCT 312
                | | | ||         |||||||||||||||||||||||||| |  ||| ||||
Db         420 GAAACAAGGTTTCCCCATGTTGCCCAGGCTGGTCTTGAACTCCTGGGCACAGGGGATCTG 361

Qy         313 CCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCT 363
               ||||| | | ||||||||    ||  || ||||||||    ||||||| ||
Db         360 CCTGCCTGGGCCTCCCAAAGTGCTGGGATTACAGATGGGAGCCACCACACT 310


Search completed: August 30, 2005, 05:17:39
Job time : 138 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 03:59:23 ; Search time 462 Seconds
(without alignments)
5169.324 Million cell updates/sec

Title: US-08-731-499-9_COPY_10001_10365
Perfect score: 365
Sequence: 1 TTTGTGGTCCTCCAAGGCTT.....GATGCACTCCACCACGCTTG 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7331713 seqs, 3271544945 residues

Total number of hits satisfying chosen parameters: 14663426

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	365	100.0	10365	8	US-08-731-499-9	Sequence 9, Appli
2	365	100.0	36022	13	US-10-087-192-1708	Sequence 1708, Ap
3	147.6	40.4	301692	17	US-10-428-487-11	Sequence 11, Appl
4	147.6	40.4	310268	19	US-10-367-094-195	Sequence 195, App
5	147.2	40.3	414295	20	US-10-719-993-6876	Sequence 6876, Ap
6	146.8	40.2	209822	21	US-10-741-600-17560	Sequence 17560, A
7	145.8	39.9	430442	20	US-10-417-375-128	Sequence 128, App
c 8	145.4	39.8	437	13	US-10-027-632-46474	Sequence 46474, A
c 9	145.4	39.8	437	13	US-10-027-632-46475	Sequence 46475, A
c 10	145.4	39.8	437	17	US-10-027-632-46474	Sequence 46474, A
c 11	145.4	39.8	437	17	US-10-027-632-46475	Sequence 46475, A
12	145	39.7	10735	20	US-10-483-241-5	Sequence 5, Appli
13	144.6	39.6	399	19	US-10-674-124A-25157	Sequence 25157, A
c 14	143.8	39.4	1980090	20	US-10-719-993-6815	Sequence 6815, Ap
c 15	143.8	39.4	1980090	21	US-10-741-600-17676	Sequence 17676, A
c 16	143.6	39.3	11216	10	US-09-764-872-490	Sequence 490, App
17	142.6	39.1	155937	20	US-10-723-860-2208	Sequence 2208, Ap
c 18	141.8	38.8	3095	13	US-10-027-632-113781	Sequence 113781,
c 19	141.8	38.8	3095	17	US-10-027-632-113781	Sequence 113781,
c 20	141.6	38.8	160274	22	US-10-893-315-140	Sequence 140, App
c 21	141.6	38.8	160300	22	US-10-893-315-151	Sequence 151, App
c 22	140.8	38.6	607	13	US-10-027-632-188264	Sequence 188264,
c 23	140.8	38.6	607	17	US-10-027-632-188264	Sequence 188264,
24	140.8	38.6	1980090	20	US-10-719-993-6815	Sequence 6815, Ap
25	140.8	38.6	1980090	21	US-10-741-600-17676	Sequence 17676, A
c 26	140.6	38.5	827	13	US-10-027-632-157392	Sequence 157392,
c 27	140.6	38.5	827	13	US-10-027-632-157393	Sequence 157393,
c 28	140.6	38.5	827	17	US-10-027-632-157392	Sequence 157392,
c 29	140.6	38.5	827	17	US-10-027-632-157393	Sequence 157393,
c 30	140.6	38.5	51917	21	US-10-741-600-17758	Sequence 17758, A
31	140.4	38.5	130349	21	US-10-741-600-17619	Sequence 17619, A
32	140.4	38.5	138363	19	US-10-367-094-117	Sequence 117, App
33	140.2	38.4	150573	22	US-10-981-277-56	Sequence 56, Appl
c 34	140.2	38.4	169659	19	US-10-322-696-70	Sequence 70, Appl
35	140	38.4	14346	17	US-10-074-024-505	Sequence 505, App
c 36	140	38.4	34739	13	US-10-087-192-1846	Sequence 1846, Ap
c 37	140	38.4	546025	20	US-10-719-993-6862	Sequence 6862, Ap
c 38	139.8	38.3	5230	15	US-10-213-948-8	Sequence 8, Appli
39	139.6	38.2	26928	17	US-10-374-979-6	Sequence 6, Appli
40	139.6	38.2	26928	18	US-10-182-936A-6	Sequence 6, Appli
41	139.6	38.2	26928	19	US-10-731-739-6	Sequence 6, Appli
42	139.6	38.2	26928	20	US-10-477-238A-6	Sequence 6, Appli
43	139.6	38.2	26928	20	US-10-680-287A-6	Sequence 6, Appli
44	139.6	38.2	26928	21	US-10-477-173-6	Sequence 6, Appli
45	139.6	38.2	26928	22	US-10-834-377-6	Sequence 6, Appli

ALIGNMENTS

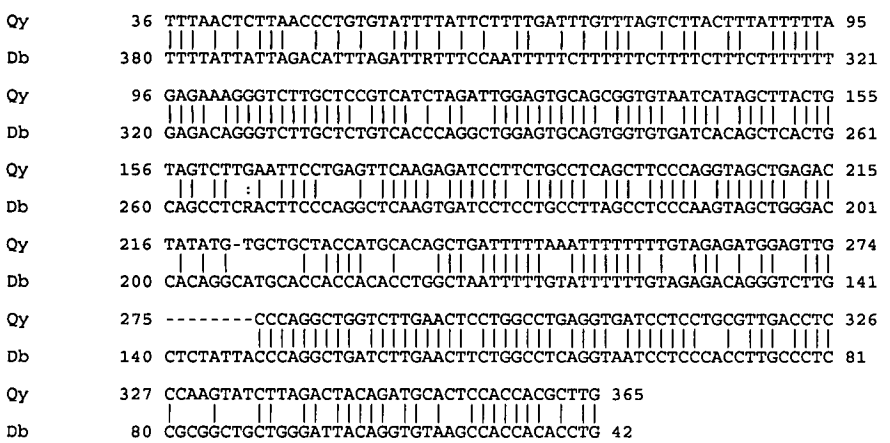
```
RESULT 1
US-08-731-499-9
; Sequence 9, Application US/08731499
; Publication No. US20030148270A1
;  GENERAL INFORMATION:
;    APPLICANT:  GRAY, Joe W.
;    APPLICANT:  COLLINS, Colin
;    APPLICANT:  HWANG, Soo-In
;    APPLICANT:  GODFREY, Tony
;    APPLICANT:  KOWBEL, David
;    APPLICANT:  ROMMENS, Johanna
;    TITLE OF INVENTION:  GENES FROM THE 20q13 AMPLICON AND THEIR
;    TITLE OF INVENTION:  USES
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/731,499
;      FILING DATE:  16-OCT-1996
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/680,395
;      FILING DATE:  15-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hunter, Tom
;      REGISTRATION NUMBER:  38,498
;      REFERENCE/DOCKET NUMBER:  2307O-068910
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10365 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..10365
;      OTHER INFORMATION:  /note= "Genomic Sequence Encoding
;      OTHER INFORMATION:  ZABC1"
US-08-731-499-9

  Query Match             100.0%;  Score 365;  DB 8;  Length 10365;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-92;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10001 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 10060

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10061 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 10120

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 10180

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 10240

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 10300

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 10360

Qy        361 GCTTG 365
              |||||
Db      10361 GCTTG 10365


RESULT 2
US-10-087-192-1708
; Sequence 1708, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1708
;   LENGTH: 36022
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-087-192-1708

  Query Match             100.0%;  Score 365;  DB 13;  Length 36022;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-91;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17531 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 17590

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17591 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 17650

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17651 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGA 17710

Qy        181 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17711 GATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGC 17770

Qy        241 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17771 TGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTGGCC 17830

Qy        301 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17831 TGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCAC 17890

Qy        361 GCTTG 365
              |||||
Db      17891 GCTTG 17895


RESULT 3
US-10-428-487-11
; Sequence 11, Application US/10428487
; Publication No. US20040006780A1
; GENERAL INFORMATION:
;  APPLICANT: RASTELLI, LUCA K.
;  APPLICANT:  GERBER, HANS-PETER
;  TITLE OF INVENTION: VEGF-MODULATED GENES AND METHODS EMPLOYING THEM
;  FILE REFERENCE: 09800080-0103
;  CURRENT APPLICATION NUMBER: US/10/428,487
;  CURRENT FILING DATE:  2003-05-02
;  PRIOR APPLICATION NUMBER: 09/815,153
;  PRIOR FILING DATE: 2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,201
;  PRIOR FILING DATE: 2000-03-22
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
;   LENGTH: 301692
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-428-487-11

  Query Match             40.4%;  Score 147.6;  DB 17;  Length 301692;
  Best Local Similarity   65.7%;  Pred. No. 4.1e-30;
  Matches  234;  Conservative    0;  Mismatches  114;  Indels    8;  Gaps    1;

Qy         18 CTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTT 77
              ||  | ||  |  | |||    |||||  |  |   |   |    |  |  | ||| |||
Db      73300 CTCTCATAGCATTATGGAAACAACTCGAATACAACTACGTCAATTATTCATCTTTTTTTT 73359

Qy         78 AGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGG 137
```

```
Db      73360 TTTTTTTTTTTTTTTTTTGAGACAGAGTCCCACTCCATCAACCAGGCTGGAGTGCAGTGG 73419

Qy        138 TGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGC 197
              || | || | ||| |||| | ||||     ||||| | ||||| ||| ||  |||||||||
Db      73420 TGCAGTCTTGGCTCACTGCATTCTTTGCATCCTGGGCTCAAGTGATTCTCGTGCCTCAGC 73479

Qy        198 TTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCTGATTTTTAAATTTTTT 257
              ||||| ||||||||||| || | ||||     ||||| | ||| | ||||||  ||||||
Db      73480 CTCCCAAGTAGCTGAGATTACAGGTGCCCACACCATTCTCAGATAATTTTTGTATTTTTA 73539

Qy        258 TTGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGAT 309
              |  |||   ||         |||| ||||||||||||| |||||||||  ||| | |||||
Db      73540 CTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAAGTGAT 73599

Qy        310 CCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
              || ||||  ||| ||||||||    ||  || ||||| |||   ||||   || ||
Db      73600 CCACCTGTCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGCGAACCACTGTGCCTG 73655


RESULT 4
US-10-367-094-195
; Sequence 195, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001500
;  CURRENT APPLICATION NUMBER: US/10/367,094
;  CURRENT FILING DATE:  2003-02-14
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 195
;   LENGTH: 310268
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-367-094-195

  Query Match             40.4%;  Score 147.6;  DB 19;  Length 310268;
  Best Local Similarity   65.7%;  Pred. No. 4.1e-30;
  Matches  234;  Conservative    0;  Mismatches  114;  Indels    8;  Gaps    1;

Qy         18 CTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTT 77
              ||  | || |  | |||    |||||  |  |   |    |     |  |  | ||| |||
Db      77830 CTCTCATAGCATTATGGAAACAACTCGAATACAACTACGTCAATTATTCATCTTTTTTTT 77889

Qy         78 AGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGG 137
              | ||  ||| ||||| |||| || |||    |||| ||| | ||  |||||||||| ||
Db      77890 TTTTTTTTTTTTTTTTTTGAGACAGAGTCCCACTCCATCAACCAGGCTGGAGTGCAGTGG 77949

Qy        138 TGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGC 197
              || | || | ||| |||| | ||||    ||||| | ||||| ||| ||  |||||||||
Db      77950 TGCAGTCTTGGCTCACTGCATTCTTTGCATCCTGGGCTCAAGTGATTCTCGTGCCTCAGC 78009

Qy        198 TTCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCTGATTTTTAAATTTTTT 257
              ||||| ||||||||||| || | ||||     ||||| | ||| | ||||||  ||||||
Db      78010 CTCCCAAGTAGCTGAGATTACAGGTGCCCACACCATTCTCAGATAATTTTTGTATTTTTA 78069

Qy        258 TTGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGAT 309
              |  |||   ||         |||| ||||||||||||| |||||||||  ||| | |||||
Db      78070 CTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAAGTGAT 78129

Qy        310 CCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
              || ||||  ||| ||||||||    ||  || ||||| |||   ||||   || ||
Db      78130 CCACCTGTCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGCGAACCACTGTGCCTG 78185


RESULT 5
US-10-719-993-6876
; Sequence 6876, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6876
;   LENGTH: 414295
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(414295)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-719-993-6876

  Query Match             40.3%;  Score 147.2;  DB 20;  Length 414295;
  Best Local Similarity   70.6%;  Pred. No. 6e-30;
  Matches  226;  Conservative    1;  Mismatches   84;  Indels    9;  Gaps    2;

Qy         52 GTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCT 111
              ||  |  ||| | ||||| |||| |||| |   |    || :|||| |||| || ||||||||
Db      88515 GTTTTATTTTTACTTTTTATTTATTTATTTATTTATTTWTTTTTGAGACAGAGTCTTGCT 88574

Qy        112 CCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTG 171
              | |||  | |   |||||||||| |||| | || |  || |||| |  ||     ||||
Db      88575 CTGTCGCCCAAGCTGGAGTGCAGTGGTGCAGTCTTGACTCACTGCAACCTCCGCCTCCTG 88634

Qy        172 AGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTAC 230
              |||||||  ||| || |||||||||| |||   |||||| ||||| | |||  ||| ||
Db      88635 GGTTCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGTGTGTGCCAC 88694

Qy        231 CATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG--------AGTTGCCCAGGCT 282
              |||||    ||| ||||||  ||||||  |  ||| | ||          ||||||||||
Db      88695 CATGCCTGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTTGCCATGTTGCCCAGGCT 88754

Qy        283 GGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACT 342
              |||||||||||||||||| |||||||| || || ||| ||||||||    ||  || |
Db      88755 GGTCTTGAACTCCTGGCCTCAGGTGATCCACCCGCCTTGGCCTCCCAAAGTGCTGGGATT 88814

Qy        343 ACAGATGCACTCCACCACGC 362
              ||||| |   |||||   ||
Db      88815 ACAGACGTGATCCACTGTGC 88834


RESULT 6
US-10-741-600-17560
; Sequence 17560, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17560
;   LENGTH: 209822
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17560

  Query Match             40.2%;  Score 146.8;  DB 21;  Length 209822;
  Best Local Similarity   69.6%;  Pred. No. 5.9e-30;
  Matches  231;  Conservative    0;  Mismatches   92;  Indels    9;  Gaps    2;

Qy         40 ACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGA 99
              |||| |    | ||| | |   | ||||  ||  ||| || ||  ||| ||||| ||||
Db     138254 ACTCATTGTATTCTGTGTGTGTGTGTTTTTTTTGTTTTTGTTTTTGTTTTTTTTTTGAGA 138313
```

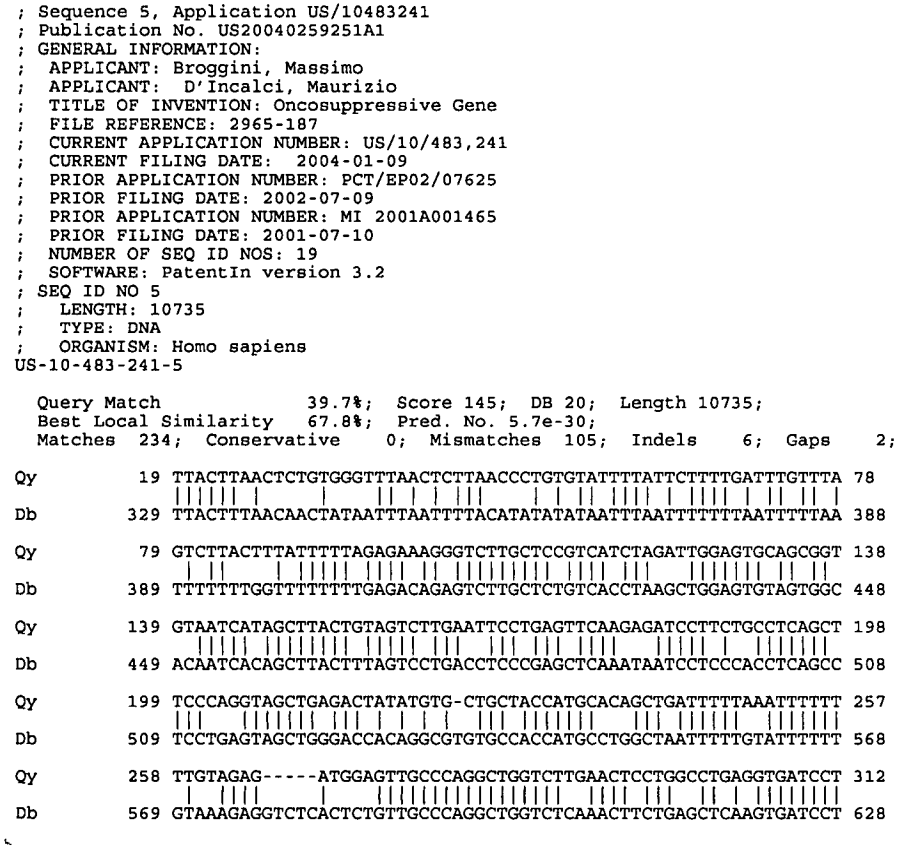
```
Qy        100 AAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGT 159
              || ||||  ||| |||| | ||  |||||||||| ||||| |||  |||| |||| |
Db     138314 CAGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAAC 138373

Qy        160 CTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATA 219
              ||  |  ||||| |||||||  || || |||||| ||| ||||| ||||||| |||||||
Db     138374 CTCCACCTCCTGGGTTCAAGCAATTCTCCTGCCTTAGCCTCCCAAGTAGCTGGGACTATA 138433

Qy        220 TGTGC-TGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG--------A 270
               | ||  || |||| || ||||| ||||||  ||||||  |  ||| | ||
Db     138434 GGCGCACGCCACCACGCCCAGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTTACCAC 138493

Qy        271 GTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAA 330
              |||| |||| | |||||||||||||| ||| ||||||||||||  | | | ||||||||
Db     138494 GTTGGCCAGACCGGTCTTGAACTCCTGACCTCAGGTGATCCTCCCACCTCGGCCTCCCAA 138553

Qy        331 GTATCTTAGACTACAGATGCACTCCACCACGC 362
              |   ||  || ||||| ||    |||||  ||
Db     138554 GGTGCTGGGATTACAGGTGTGAGCCACCGTGC 138585


RESULT 7
US-10-417-375-128
; Sequence 128, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/417,375
;  CURRENT FILING DATE:  2003-04-15
;  NUMBER OF SEQ ID NOS: 176
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 128
;   LENGTH: 430442
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-417-375-128

  Query Match             39.9%;  Score 145.8;  DB 20;  Length 430442;
  Best Local Similarity   67.3%;  Pred. No. 1.5e-29;
  Matches  239;  Conservative    0;  Mismatches  107;  Indels    9;  Gaps    2;

Qy         19 TTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTA 78
              | || ||  ||| | | | | | | |       | |   |||  ||  |||  ||| |||
Db     221592 TCACCTATTTCTTTTGATGTTATTTTATTTTATTTTATTTTATTTTATTTATTTTATTTT 221651

Qy         79 GTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGT 138
              | ||| |||||||||  |||| || |||| |||  |||| | ||  |||||||||| ||
Db     221652 ATTTTATTTTATTTTTTGAGACAGAGTCTGGCTGTGTCACCCAGGCTGGAGTGCAGTGGC 221711

Qy        139 GTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCT 198
                |||   |||  ||| |  ||      ||||| ||||||| ||| || ||||||||||
Db     221712 ACGATCTCGGCTCGCTGCAACCTCTGCCTCCTGGGTTCAAGTGATTCTCCTGCCTCAGCC 221771

Qy        199 TCCCAGGTAGCTGAGACTATATGTGCTGCTACCATGCACAGCTGATTTTTAAATTTTTTT 258
              |||||||||||| || || | |  || | |||| || ||||| ||||||  ||||||
Db     221772 TCCCAGGTAGCTGGGATTACAAGCACT-CCACCACGCCCAGCTAATTTTTGTATTTTTAG 221830

Qy        259 TGTAGAGATGG--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATC 310
              || |||   ||          |||| |||||||||| | ||||||||||||  | ||||||
Db     221831 TGGAGATGGGGTTTCACTATGTTGGCCAGGCTGGTTTGGAACTCCTGGCCTCACGTGATC 221890

Qy        311 CTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
              | ||||| ||| ||||||||     ||  || ||||| ||    ||||| | | ||
Db     221891 CACCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCGCACCTG 221945


RESULT 8
US-10-027-632-46474/c
; Sequence 46474, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46474
;   LENGTH: 437
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-46474

  Query Match             39.8%;  Score 145.4;  DB 13;  Length 437;
  Best Local Similarity   68.4%;  Pred. No. 1.2e-30;
  Matches  232;  Conservative    1;  Mismatches   97;  Indels    9;  Gaps    2;

Qy         36 TTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTA 95
              ||| | | |||  | | |  ||| | | |  ||  | | |||  | ||  ||  |||||
Db        380 TTTTATTATTAGACATTTAGATTRTTTCCAATTTTTCTTTTTTCTTTTCTTTCTTTTTTT 321

Qy         96 GAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTG 155
              |||| ||||||||||||| |||| | ||  |||||||||| ||||| |||| |||| ||||
Db        320 GAGACAGGGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCACAGCTCACTG 261

Qy        156 TAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGAC 215
              || || :| ||||    | ||||| |||||| |||||| ||| ||||| ||||||| |||
Db        260 CAGCCTCRACTTCCCAGGCTCAAGTGATCCTCCTGCCTTAGCCTCCCAAGTAGCTGGGAC 201

Qy        216 TATATG-TGCTGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTG 274
              | | |    | ||||  |   ||| |||||  |||||||  |  ||| | ||  |||
Db        200 CACAGGCATGCACCACCACACCTGGCTAATTTTTGTATTTTTTGTAGAGACAGGGTCTTG 141

Qy        275 --------CCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTC 326
                     ||||||||| ||||||||| ||||||| |||| |||||||  | ||| ||||
Db        140 CTCTATTACCCAGGCTGATCTTGAACTTCTGGCCTCAGGTAATCCTCCCACCTTGCCCTC 81

Qy        327 CCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
              |   |   ||  || ||||| || |  |||||| | ||
Db         80 CGCGGCTGCTGGGATTACAGGTGTAAGCCACCACACCTG 42


RESULT 9
US-10-027-632-46475/c
; Sequence 46475, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
```

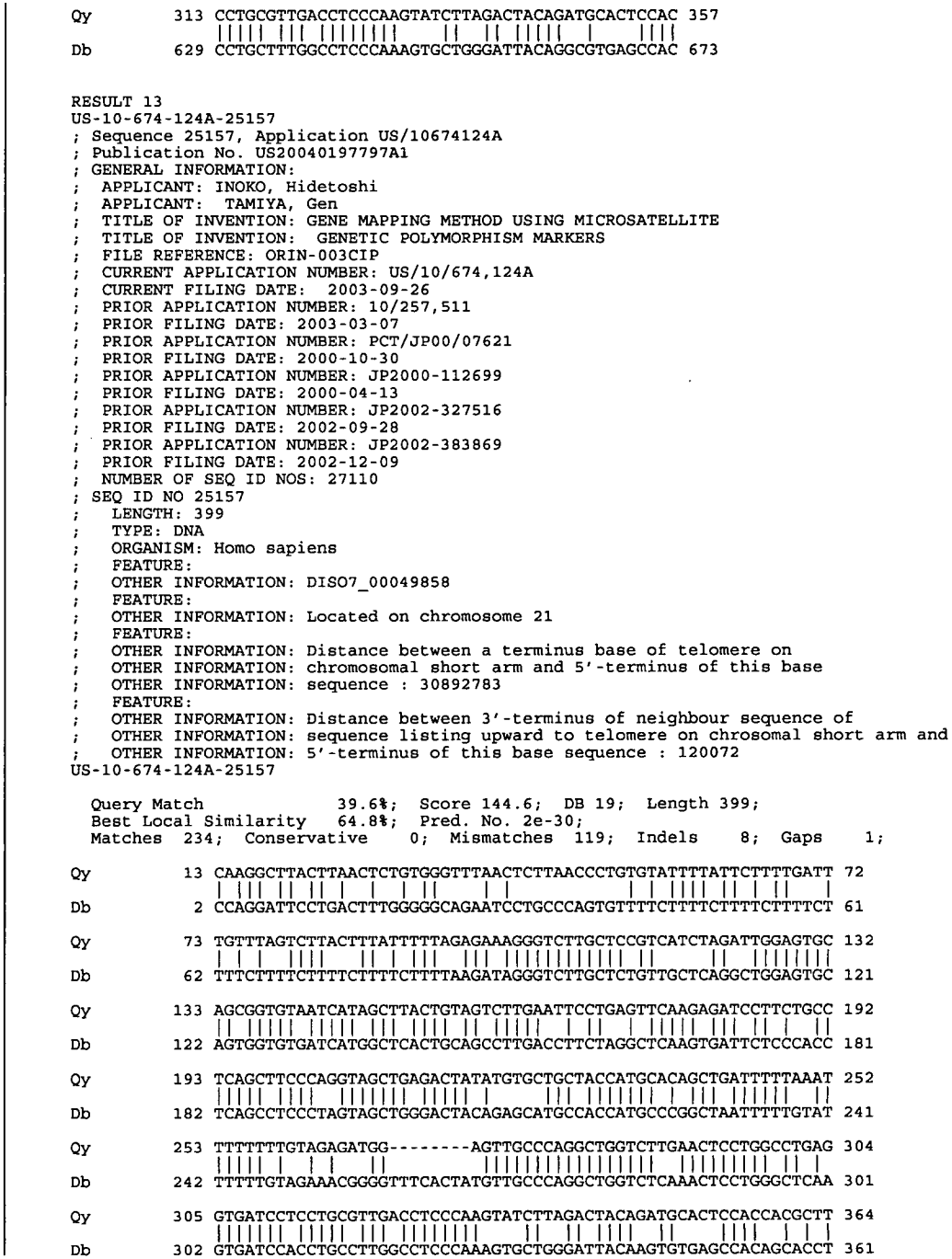
```
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46475
;   LENGTH: 437
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-46475

  Query Match             39.8%;  Score 145.4;  DB 13;  Length 437;
  Best Local Similarity   68.4%;  Pred. No. 1.2e-30;
  Matches  232;  Conservative    1;  Mismatches   97;  Indels    9;  Gaps    2;

Qy         36 TTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTA 95
              ||| | | |||  | | |  ||| | | |  ||  | | |||  | ||  ||  |||||
Db        380 TTTTATTATTAGACATTTAGATTRTTTCCAATTTTTCTTTTTTCTTTTCTTTCTTTTTTT 321

Qy         96 GAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTG 155
              |||| ||||||||||| |||| | ||  |||||||||| ||||| |||| |||| ||||
Db        320 GAGACAGGGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCACAGCTCACTG 261

Qy        156 TAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGAC 215
              || || :| ||||   | ||||| |||||| |||||| ||| ||||| ||||||| |||
Db        260 CAGCCTCRACTTCCCAGGCTCAAGTGATCCTCCTGCCTTAGCCTCCCAAGTAGCTGGGAC 201

Qy        216 TATATG-TGCTGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTG 274
              | | |      | ||||  |   ||| ||||||  ||||||| |  ||| | ||  |||
Db        200 CACAGGCATGCACCACCACACCTGGCTAATTTTTGTATTTTTTGTAGAGACAGGGTCTTG 141

Qy        275 --------CCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTC 326
                    ||||||||| |||||||||| ||||||| |||| ||||||| | ||| ||||
Db        140 CTCTATTACCCAGGCTGATCTTGAACTTCTGGCCTCAGGTAATCCTCCCACCTTGCCCTC 81

Qy        327 CCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
              |   |   ||  || ||||| || |  ||||||| | ||
Db         80 CGCGGCTGCTGGGATTACAGGTGTAAGCCACCACACCTG 42


RESULT 10
US-10-027-632-46474/c
; Sequence 46474, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46474
;   LENGTH: 437
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-46474

  Query Match             39.8%;  Score 145.4;  DB 17;  Length 437;
  Best Local Similarity   68.4%;  Pred. No. 1.2e-30;
  Matches  232;  Conservative    1;  Mismatches   97;  Indels    9;  Gaps    2;

Qy         36 TTTAACTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTA 95
              ||| | | |||  | | |  ||| | | |  ||  | | |||  | ||  ||  |||||
Db        380 TTTTATTATTAGACATTTAGATTRTTTCCAATTTTTCTTTTTTCTTTTCTTTCTTTTTTT 321

Qy         96 GAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTG 155
              |||| ||||||||||| |||| | ||  |||||||||| ||||| |||| |||| ||||
Db        320 GAGACAGGGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCACAGCTCACTG 261

Qy        156 TAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGAC 215
              || || :| ||||   | ||||| |||||| |||||| ||| ||||| ||||||| |||
Db        260 CAGCCTCRACTTCCCAGGCTCAAGTGATCCTCCTGCCTTAGCCTCCCAAGTAGCTGGGAC 201

Qy        216 TATATG-TGCTGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTG 274
              | | |      | ||||  |   ||| ||||||  ||||||| |  ||| | ||  |||
Db        200 CACAGGCATGCACCACCACACCTGGCTAATTTTTGTATTTTTTGTAGAGACAGGGTCTTG 141

Qy        275 --------CCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTC 326
                    ||||||||| |||||||||| ||||||| |||| ||||||| | ||| ||||
Db        140 CTCTATTACCCAGGCTGATCTTGAACTTCTGGCCTCAGGTAATCCTCCCACCTTGCCCTC 81

Qy        327 CCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
              |   |   ||  || ||||| || |  ||||||| | ||
Db         80 CGCGGCTGCTGGGATTACAGGTGTAAGCCACCACACCTG 42


RESULT 11
US-10-027-632-46475/c
; Sequence 46475, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 46475
;   LENGTH: 437
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-46475

  Query Match             39.8%;  Score 145.4;  DB 17;  Length 437;
  Best Local Similarity   68.4%;  Pred. No. 1.2e-30;
  Matches  232;  Conservative    1;  Mismatches   97;  Indels    9;  Gaps    2;
```



RESULT 12
US-10-483-241-5
; Sequence 5, Application US/10483241
; Publication No. US20040259251A1
; GENERAL INFORMATION:
; APPLICANT: Broggini, Massimo
; APPLICANT: D'Incalci, Maurizio
; TITLE OF INVENTION: Oncosuppressive Gene
; FILE REFERENCE: 2965-187
; CURRENT APPLICATION NUMBER: US/10/483,241
; CURRENT FILING DATE: 2004-01-09
; PRIOR APPLICATION NUMBER: PCT/EP02/07625
; PRIOR FILING DATE: 2002-07-09
; PRIOR APPLICATION NUMBER: MI 2001A001465
; PRIOR FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 10735
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-483-241-5

Query Match 39.7%; Score 145; DB 20; Length 10735;
Best Local Similarity 67.8%; Pred. No. 5.7e-30;
Matches 234; Conservative 0; Mismatches 105; Indels 6; Gaps 2;





RESULT 13
US-10-674-124A-25157
; Sequence 25157, Application US/10674124A
; Publication No. US20040197797A1
; GENERAL INFORMATION:
; APPLICANT: INOKO, Hidetoshi
; APPLICANT: TAMIYA, Gen
; TITLE OF INVENTION: GENE MAPPING METHOD USING MICROSATELLITE
; TITLE OF INVENTION: GENETIC POLYMORPHISM MARKERS
; FILE REFERENCE: ORIN-003CIP
; CURRENT APPLICATION NUMBER: US/10/674,124A
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: 10/257,511
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: PCT/JP00/07621
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: JP2000-112699
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: JP2002-327516
; PRIOR FILING DATE: 2002-09-28
; PRIOR APPLICATION NUMBER: JP2002-383869
; PRIOR FILING DATE: 2002-12-09
; NUMBER OF SEQ ID NOS: 27110
; SEQ ID NO 25157
; LENGTH: 399
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: DISO7_00049858
; FEATURE:
; OTHER INFORMATION: Located on chromosome 21
; FEATURE:
; OTHER INFORMATION: Distance between a terminus base of telomere on
; OTHER INFORMATION: chromosomal short arm and 5'-terminus of this base
; OTHER INFORMATION: sequence : 30892783
; FEATURE:
; OTHER INFORMATION: Distance between 3'-terminus of neighbour sequence of
; OTHER INFORMATION: sequence listing upward to telomere on chrosomal short arm and
; OTHER INFORMATION: 5'-terminus of this base sequence : 120072
US-10-674-124A-25157

Query Match 39.6%; Score 144.6; DB 19; Length 399;
Best Local Similarity 64.8%; Pred. No. 2e-30;
Matches 234; Conservative 0; Mismatches 119; Indels 8; Gaps 1;

```
Qy        365 G 365
              |
Db        362 G 362


RESULT 14
US-10-719-993-6815/c
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6815
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6815

  Query Match             39.4%;  Score 143.8;  DB 20;  Length 1980090;
  Best Local Similarity   69.8%;  Pred. No. 1e-28;
  Matches  227;  Conservative    0;  Mismatches   87;  Indels   11;  Gaps    2;

Qy         51 TGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGC 110
              ||| |  |||| | ||||  ||| |||  | ||  ||| ||||  || | ||||||| ||
Db     180290 TGTTTCCTTTACTTTTTTTCTTTCTTTCTTTTTTTTTTTTTTTTGGGAAACAGGGTCTGGC 180231

Qy        111 TCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCT 170
              || |||| | ||  |||||||||| ||||| |||  |||| |||| |  ||  |  ||||
Db     180230 TCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACCTCCACCTCCT 180171

Qy        171 GAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCT-GCTA 229
              | | |||||  ||||| |  |||||||  ||   ||||||| ||||||| ||| |  | |
Db     180170 GGGCTCAAGCTATCCTCCCACCTCAGCACCCTGAGTAGCTGGGACTATAGGTGTTCACCA 180111

Qy        230 CCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG----------AGTTGCCCAG 279
              ||| || ||||| ||||||  |||||||| ||||||||||          |||||||||
Db     180110 CCAAGCCCAGCTAATTTTTGTATTTTTTTGGTAGAGATGGGGTTTTGCCATGTTGCCCAG 180051

Qy        280 GCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAG 339
              |||| |||||||||||||| || |  ||||||     | |   ||||||||   |||  |
Db     180050 GCTGATCTTGAACTCCTGGGCTCAAATGATCCATTCACCTCAGCCTCCCAAAGTTCTGGG 179991

Qy        340 ACTACAGATGCACTCCACCACGCTT 364
              | ||||| ||    ||||||| | |
Db     179990 ATTACAGGTGTGAGCCACCACTCCT 179966


RESULT 15
US-10-741-600-17676/c
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17676
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17676

  Query Match             39.4%;  Score 143.8;  DB 21;  Length 1980090;
  Best Local Similarity   69.8%;  Pred. No. 1e-28;
  Matches  227;  Conservative    0;  Mismatches   87;  Indels   11;  Gaps    2;

Qy         51 TGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGC 110
              ||| |  |||| | ||||  ||| |||  | ||  ||| ||||  || | ||||||| ||
Db     180290 TGTTTCCTTTACTTTTTTTCTTTCTTTCTTTTTTTTTTTTTTTTGGGAAACAGGGTCTGGC 180231

Qy        111 TCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCT 170
              || |||| | ||  |||||||||| ||||| |||  |||| |||| |  ||  |  ||||
Db     180230 TCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACCTCCACCTCCT 180171

Qy        171 GAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCT-GCTA 229
              | | |||||  ||||| |  |||||||  ||   ||||||| ||||||| ||| |  | |
Db     180170 GGGCTCAAGCTATCCTCCCACCTCAGCACCCTGAGTAGCTGGGACTATAGGTGTTCACCA 180111

Qy        230 CCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG----------AGTTGCCCAG 279
              ||| || ||||| ||||||  |||||||| ||||||||||          |||||||||
Db     180110 CCAAGCCCAGCTAATTTTTGTATTTTTTTGGTAGAGATGGGGTTTTGCCATGTTGCCCAG 180051

Qy        280 GCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAG 339
              |||| |||||||||||||| || |  ||||||     | |   ||||||||   |||  |
Db     180050 GCTGATCTTGAACTCCTGGGCTCAAATGATCCATTCACCTCAGCCTCCCAAAGTTCTGGG 179991

Qy        340 ACTACAGATGCACTCCACCACGCTT 364
              | ||||| ||    ||||||| | |
Db     179990 ATTACAGGTGTGAGCCACCACTCCT 179966


Search completed: August 30, 2005, 05:25:32
Job time : 469 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: August 30, 2005, 02:54:04 ; Search time 2193 Seconds
(without alignments)
6335.366 Million cell updates/sec

Title: US-08-731-499-9_COPY_10001_10365
Perfect score: 365
Sequence: 1 TTTGTGGTCCTCCAAGGCTT.....GATGCACTCCACCACGCTTG 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	166	45.5	511	8	AQ512276	AQ512276 HS_5077_A
c	2	142	38.9	751	8	BZ612069	BZ612069 WHADP28TR
	3	141.6	38.8	2225	3	BC033179	BC033179 Homo sapi
	4	140.2	38.4	535	8	AQ213795	AQ213795 HS_2178_B
c	5	140	38.4	516	5	BU076629	BU076629 im50c03.y
	6	139.6	38.2	1000	4	BM452708	BM452708 AGENCOURT
	7	139.4	38.2	485	7	CN263690	CN263690 170004241
	8	139.4	38.2	678	5	BX481790	BX481790 DKFZp686K
	9	139.4	38.2	788	7	CK000205	CK000205 AGENCOURT
	10	139.4	38.2	837	6	CD656882	CD656882 AGENCOURT
	11	139.2	38.1	422	7	H73550	H73550 ys10h07.r1
c	12	138.8	38.0	384	2	AW884394	AW884394 QV3-OT006
c	13	138.8	38.0	2198	3	HSM806754	BX641053 Homo sapi
	14	138.6	38.0	8056	3	CR749384	CR749384 Homo sapi
	15	138.4	37.9	360	1	AI332671	AI332671 qq33f08.x
c	16	138.4	37.9	509	8	AQ347397	AQ347397 RPCI11-10
	17	138.4	37.9	741	1	AI679747	AI679747 tu76c03.x
c	18	138	37.8	482	8	AQ133619	AQ133619 HS_3047_A
	19	138	37.8	677	9	AG094571	AG094571 Pan trogl
	20	137.8	37.8	561	5	BX475065	BX475065 DKFZp686F
	21	137.8	37.8	606	6	CA748785	CA748785 UI-H-FT1-
c	22	137.6	37.7	598	6	CD242461	CD242461 AGENCOURT
	23	137.6	37.7	629	8	AQ422183	AQ422183 RPCI-11-2
	24	137.4	37.6	329	1	AI358712	AI358712 qx14c02.x
	25	137.2	37.6	315	1	AI014798	AI014798 ot79g03.s
	26	137.2	37.6	558	6	CB123312	CB123312 K-EST0171
	27	137.2	37.6	562	6	CA398239	CA398239 cs99g11.x
c	28	137.2	37.6	1063	4	BM456792	BM456792 AGENCOURT
	29	137	37.5	440	1	AA129746	AA129746 zn91g07.s
c	30	136.4	37.4	538	6	CD691041	CD691041 EST7564 h
	31	136	37.3	368	1	AI590042	AI590042 tr75c02.x
c	32	135.8	37.2	884	4	BM044361	BM044361 603621675
c	33	135.6	37.2	699	4	BG563044	BG563044 602581828
c	34	135.6	37.2	702	9	AG145061	AG145061 Pan trogl
	35	135.2	37.0	627	8	AQ315924	AQ315924 RPCI11-96
c	36	135	37.0	629	8	AQ320405	AQ320405 RPCI11-10
	37	135	37.0	672	5	BU929449	BU929449 AGENCOURT
c	38	135	37.0	4650	3	BC033208	BC033208 Homo sapi
	39	134.8	36.9	421	7	CN289800	CN289800 170004243
	40	134.8	36.9	439	8	B32496	B32496 HS-1015-B1-
	41	134.8	36.9	501	2	BF769916	BF769916 RC3-IT001
	42	134.8	36.9	584	8	AQ111520	AQ111520 CIT-HSP-2
	43	134.8	36.9	640	8	AQ385850	AQ385850 RPCI11-15
c	44	134.8	36.9	877	8	AQ749458	AQ749458 HS_5574_A
c	45	134.4	36.8	338	2	AW023111	AW023111 df49d08.y

ALIGNMENTS

RESULT 1
AQ512276/c
LOCUS AQ512276 511 bp DNA linear GSS 05-MAY-1999
DEFINITION HS_5077_A2_A10_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=653 Col=20 Row=A, genomic survey sequence.
ACCESSION AQ512276
VERSION AQ512276.1 GI:4742829
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 511)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 653 row: A column: 20
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 511.
FEATURES Location/Qualifiers
source 1. .511
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=653 Col=20 Row=A"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

ORIGIN

Query Match 45.5%; Score 166; DB 8; Length 511;
Best Local Similarity 100.0%; Pred. No. 4.8e-24;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 TTTGTGGTCCTCCAAGGCTTACTTAACTCTGTGGGTTTAACTCTTAACCCTGTGTATTTT 109

Qy         61 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        108 ATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCT 49

Qy        121 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAAT 166
              ||||||||||||||||||||||||||||||||||||||||||||||
Db         48 AGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAAT 3
```

RESULT 2
BZ612069/c
LOCUS BZ612069 751 bp DNA linear GSS 08-JUN-2003
DEFINITION WHADP28TR Human MCF7 breast cancer cell line library (MCF7_1) Homo sapiens genomic clone MCF7_1-22F8, genomic survey sequence.
ACCESSION BZ612069
VERSION BZ612069.1 GI:31520630
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 751)
AUTHORS Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q., Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P., Gray,J.W. and Collins,C.
TITLE End-sequence profiling: Sequence-based analysis of aberrant genomes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
MEDLINE 22709111
PUBMED 12788976
COMMENT Contact: Volik SV
Colin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolik@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .751
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="MCF7_1-22F8"
/sex="female"
/clone_lib="Human MCF7 breast cancer cell line library (MCF7_1)"
/note="Vector: pECBAC1; Site_1: HindIII; This library was constructed from MCF7 breast cancer cell line by Amplicon Express (http://www.genomex.com) using their standard procedure."

ORIGIN

Query Match 38.9%; Score 142; DB 8; Length 751;
Best Local Similarity 69.4%; Pred. No. 3.7e-19;
Matches 225; Conservative 0; Mismatches 90; Indels 9; Gaps 2;

```
Qy         48 CCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCT 107
              | ||   | ||||    ||||| ||  ||| || ||  ||| ||||| ||||  | | ||
Db        443 CTCTTCATTTTTTGGGTTTTTGTTTGTTTTTGTTTTTTTTTTTTTTTTTGAGACGGAGCCT 384

Qy        108 TGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATT 167
              ||||| |||| | ||  |||||||||| |||||||||  | ||| |||| || ||     |
Db        383 TGCTCTGTCACCCAGGCTGGAGTGCAGTGGTGTAATCTTGGCTCACTGCAGCCTCTGCCT 324

Qy        168 CCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTG-CTG 226
              || | ||||||| ||| || ||||||||||||||  |||||| |  |||| ||| ||
Db        323 CCCGGGTTCAAGTGATTCTCCTGCCTCAGCTTCCCGAGTAGCTGGGGTTATAGGTGCCTC 264

Qy        227 CTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG--------AGTTGCCCA 278
              |||||| || | ||| ||||||  ||||||  |  ||| | ||         | || |||
Db        263 CTACCACGCCCGGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGCTGGCCA 204

Qy        279 GGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTA 338
              ||||||||||||||||   ||| ||||||||  ||  | | | ||||||||    ||
Db        203 GGCTGGTCTTGAACTCCCAACCTCAGGTGATCTGCCCACCTCGGCCTCCCAAAGTGCTGG 144

Qy        339 GACTACAGATGCACTCCACCACGC 362
              || |||||  |    ||||| |||
Db        143 GATTACAGGCGTGAGCCACCGCGC 120
```

RESULT 3
BC033179
LOCUS BC033179 2225 bp mRNA linear HTC 01-APR-2004
DEFINITION Homo sapiens ATPase, Class I, type 8B, member 3, mRNA (cDNA clone IMAGE:4581020), with apparent retained intron.
ACCESSION BC033179
VERSION BC033179.1 GI:21619903
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2225)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932
REFERENCE 2 (bases 1 to 2225)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield, Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu, Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska, Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 43 Row: f Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis
This clone has the following problem: retained intron.

FEATURES Location/Qualifiers
source 1. .2225
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4581020"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH_MGC_16"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN

Query Match 38.8%; Score 141.6; DB 3; Length 2225;
Best Local Similarity 71.2%; Pred. No. 3.7e-19;
Matches 218; Conservative 0; Mismatches 79; Indels 9; Gaps 2;

```
Qy         66 TTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATT 125
Db       1782 TTTTAAACAGTGTGTATTTATTTATTTTTTGAGATGGGGTCTTGCTCTGCCGCCCAGGCT 1841

Qy        126 GGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCC 185
Db       1842 GGAATGCAGTGGTGCAATCATAGCTTATTGCAGCCTCGAATTCCTGGGCTCAAGCAATCC 1901

Qy        186 TTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGAT 244
Db       1902 TCCCACTTCAGCTTCCCAAGTAGCCGGGACTATAGGAGAGTGCCACCTTACCCAGCTTAT 1961

Qy        245 TTTTAAATTTTTTTTGTAGAGATGGA--------GTTGCCCAGGCTGGTCTTGAACTCCT 296
Db       1962 TTTTGTATTTTTTGTCAAGACAGGGAATCCCTATGTTGCCCAGGCTGGTCTTGAACTCCT 2021

Qy        297 GGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCA 356
Db       2022 GGGCTTAAGCGATCCGCCTGCCTCCGCTTTTCAAAGCACTGGAATTACAGATGTGAGCCA 2081

Qy        357 CCACGC 362
Db       2082 CCACAC 2087
```

RESULT 4
AQ213795
LOCUS AQ213795 535 bp DNA linear GSS 18-SEP-1998
DEFINITION HS_2178_B2_A10_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2178 Col=20 Row=B, genomic survey sequence.
ACCESSION AQ213795
VERSION AQ213795.1 GI:3624996
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 535)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2178 row: B column: 20
Class: BAC ends
High quality sequence stop: 535.

FEATURES Location/Qualifiers
source 1. .535
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=2178 Col=20 Row=B"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 38.4%; Score 140.2; DB 8; Length 535;
Best Local Similarity 70.1%; Pred. No. 9.1e-19;
Matches 216; Conservative 0; Mismatches 89; Indels 3; Gaps 2;

```
Qy         59 TTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCAT 118
Db        106 TTTCTATTTAGTTNTGTTTTGCTTTGTTTTGTGGTTTGAGACAGAGTCTTGCTGTGTTGC 165

Qy        119 CTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAA 178
Db        166 CCAGGCTGGAGTGCAATGGTGTGATCTTGGCTCACTGCAGCCTCCGCCTCCCAGGTTCAA 225

Qy        179 GAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGC-TGCTACCATGCAC 237
Db        226 GTGATTCTCCTGCCTCAGCCTACCGAGCAGCTGGGACTACAGGTGCGTGCCACCACGTCC 285

Qy        238 AGCTGATTTTTAAATTTTTTTTGTAGAGATGGAGTTGCCCAGGCTGGTCTTGAACTCCTG 297
Db        286 AGCTAATTTTT--GTATTTTTAGTAGAGACCATGTTGGCCAGGCTGGTCTCGAACTCCTG 343

Qy        298 GCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCAC 357
Db        344 ACCTCAAGTGATCTGCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCAC 403

Qy        358 CACGCTTG 365
Db        404 CACACCTG 411
```

RESULT 5
BU076629/c
LOCUS BU076629 516 bp mRNA linear EST 27-AUG-2002
DEFINITION im50c03.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6038404 5', mRNA sequence.
ACCESSION BU076629
VERSION BU076629.1 GI:22517811
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 516)

```
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,
            Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 468.
FEATURES             Location/Qualifiers
     source          1. .516
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6038404"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
ORIGIN

  Query Match             38.4%;  Score 140;  DB 5;  Length 516;
  Best Local Similarity   70.9%;  Pred. No. 1e-18;
  Matches  217;  Conservative    0;  Mismatches   80;  Indels    9;  Gaps    2;

Qy         66 TTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATT 125
              ||| |     |  || ||  |||||||| ||||  |||||||||| | |  | ||  |
Db        421 TTTTAAACAGTGTGTATTTATTTATTTTTTGAGATGGGGTCTTGCTCTGCCGCCCAGGCT 362

Qy        126 GGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCC 185
              ||| ||||| |||| |||||||||||| || || || ||||||||| | |||||  ||||
Db        361 GGAATGCAGTGGTGCAATCATAGCTTATTGCAGCCTCGAATTCCTGGGCTCAAGCAATCC 302

Qy        186 TTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTACCATGCACAGCTGAT 244
              | |  | |||||||||| ||||| | ||||| | | |  ||| ||| | | ||||| ||
Db        301 TCCCACTTCAGCTTCCCAAGTAGCCGGGACTACAGGAGAGTGCCACCTTACCCAGCTTAT 242

Qy        245 TTTTAAATTTTTTTTGTAGAGATGGA--------GTTGCCCAGGCTGGTCTTGAACTCCT 296
              ||||  ||||||| |  ||| | |||        ||||||||||||||||||||||||
Db        241 TTTTGTATTTTTTGTCAAGACAGGGAATCCCTATGTTGCCCAGGCTGGTCTTGAACTCCT 182

Qy        297 GGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCA 356
              || || | | ||||| |||||  |  | | |||    ||   | ||||||||    |||
Db        181 GGGCTTAAGCGATCCGCCTGCCTCCGCTTTTCAAAGCACTGGAATTACAGATGTGAGCCA 122

Qy        357 CCACGC 362
              |||| |
Db        121 CCACAC 116


RESULT 6
BM452708
LOCUS       BM452708               1000 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6401261 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5498573
            5', mRNA sequence.
ACCESSION   BM452708
VERSION     BM452708.1  GI:18501748
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1000)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12130  row: i  column: 06
            High quality sequence stop: 634.
FEATURES             Location/Qualifiers
     source          1. .1000
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5498573"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_85"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             38.2%;  Score 139.6;  DB 4;  Length 1000;
  Best Local Similarity   68.1%;  Pred. No. 1.1e-18;
  Matches  226;  Conservative    0;  Mismatches   99;  Indels    7;  Gaps    2;

Qy         41 CTCTTAACCCTGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAA 100
              ||  |  |  | | |||    ||   |   ||  |||  | ||   |  ||||||||||
Db        396 CTTCTTTCTGTTTCTATGAATTTGACTACTTTGTTTTTATTTTCTATATTTTTTAGAGAC 455

Qy        101 AGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTC 160
              | |||||||||| ||  || |   ||||| |||| ||||  || ||||||  |||| |  |
Db        456 AAGGTCTTGCTCTGTTGTCCAAGCTGGAGGGCAGTGGTGCGATTATAGCTCACTGCAACC 515

Qy        161 TTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATAT 220
              | ||| |||   | ||||| |||||| ||  |||||| ||||| ||||||  ||  | |
Db        516 TGGAACTCCCAGGCTCAAGTGATCCTCCTCACTCAGCCTCCCAAGTAGCTAGGATGACAG 575

Qy        221 GTGC-TGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAG------ATGGAGTT 273
              |  | ||| ||||| | ||||| |||||||||  |||||| |||||         | |||
Db        576 GGACTTGCCACCATACCCAGCTTATTTTTAAAAATTTTTTATAGAGGGCCTCGCTGTGTT 635

Qy        274 GCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTA 333
              ||||| ||| |||||||||||||||||| | |||||||||||| ||| ||||||||
Db        636 GCCCAAGCTAGTCTTGAACTCCTGGCCTCAAGTGATCCTCCTGCCTTGGCCTCCCAAAGT 695

Qy        334 TCTTAGACTACAGATGCACTCCACCACGCTTG 365
                 || ||  |||||       ||||||  || ||
Db        696 GCTGAGGTTACAGGCATGAGCCACCAAGCCTG 727


RESULT 7
CN263690
LOCUS       CN263690               485 bp    mRNA    linear   EST 16-MAY-2004
```

```
DEFINITION  17000424177371 GRN_ES Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN263690
VERSION     CN263690.1  GI:47280104
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 485   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .485
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, cell lines H1, H7, and
                     H9"
                     /clone_lib="GRN_ES"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from undifferentiated hES cell lines H1 (p32), H7 (p29),
                     and H9 (p26) maintained in feeder-free conditions"
ORIGIN

  Query Match            38.2%;  Score 139.4;  DB 7;  Length 485;
  Best Local Similarity  69.0%;  Pred. No. 1.3e-18;
  Matches  223;  Conservative    0;  Mismatches   91;  Indels    9;  Gaps    2;

Qy         52 GTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCT 111
              | || |||| || ||||| |||||||  | |    ||  ||||| |||| | |||| |||
Db         92 GGGTTTTTTTTTGTTTTGTTTTGTTTGTTTTGTTTTGTTTTTTTGAGACGGTGTCTCGCT 151

Qy        112 CCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTG 171
              | ||| | ||  |||||||||| || | ||||    ||| ||| |||||     |||||
Db        152 GCATCACCCAGGCTGGAGTGCAGTGGCGCAATCTCGCCTTCCTGCAGTCTCCGCCTCCTG 211

Qy        172 AGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACC 231
              |||||||| ||| || |||||||||| ||||| ||||||| ||||  | | ||    |||
Db        212 AGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTGCAGGCGCGCACACC 271

Qy        232 ATGCACAGCTGATTTTTAAATTTTTTTTGTAGA--------GATGGAGTTGCCCAGGCTG 283
              | || ||||| ||||||  ||||||  |  |||         ||   |||| ||||||||
Db        272 ACGCCCAGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCATCATGTTGACCAGGCTG 331

Qy        284 GTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGA-CCTCCCAAGTATCTTAGACT 342
              ||||  |||||||| | | ||||||||| || || || | ||||||||    ||  || |
Db        332 GTCTCAAACTCCTGACTTCAGGTGATCCACCCGCCTTCAGCCTCCCAACGTGCTGGGATT 391

Qy        343 ACAGATGCACTCCACCACGCTTG 365
              ||||  |    || || ||| ||
Db        392 ACAGCCGTGAGCCGCCGCGCCTG 414


RESULT 8
BX481790
LOCUS       BX481790                678 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp686K11227_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686K11227 5', mRNA sequence.
ACCESSION   BX481790
VERSION     BX481790.1  GI:31941068
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     EST (Bahr,A., Lauber,J., Mewes,H.W., Weil,B., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No s1 sequence available.
            This clone (DKFZp686K11227) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .678
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686K11227"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match            38.2%;  Score 139.4;  DB 5;  Length 678;
  Best Local Similarity  69.0%;  Pred. No. 1.3e-18;
  Matches  223;  Conservative    0;  Mismatches   91;  Indels    9;  Gaps    2;

Qy         52 GTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCT 111
              | || |||| || ||||| |||||||  | |    ||  ||||| |||| | |||| |||
Db        101 GGGTTTTTTTTTGTTTTGTTTTGTTTGTTTTGTTTTGTTTTTTTGAGACGGTGTCTCGCT 160

Qy        112 CCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTG 171
              | ||| | ||  |||||||||| || | ||||    ||| ||| |||||     |||||
Db        161 GCATCACCCAGGCTGGAGTGCAGTGGCGCAATCTCGCCTTCCTGCAGTCTCCGCCTCCTG 220

Qy        172 AGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACC 231
              |||||||| ||| || |||||||||| ||||| ||||||| ||||  | | ||    |||
Db        221 AGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTGCAGGCGCGCACACC 280

Qy        232 ATGCACAGCTGATTTTTAAATTTTTTTTGTAGA--------GATGGAGTTGCCCAGGCTG 283
              | || ||||| ||||||  ||||||  |  |||         ||   |||| ||||||||
Db        281 ACGCCCAGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCATCATGTTGACCAGGCTG 340

Qy        284 GTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGA-CCTCCCAAGTATCTTAGACT 342
              ||||  |||||||| | | ||||||||| || || || | ||||||||    ||  || |
Db        341 GTCTCAAACTCCTGACTTCAGGTGATCCACCCGCCTTCAGCCTCCCAACGTGCTGGGATT 400

Qy        343 ACAGATGCACTCCACCACGCTTG 365
              ||||  |    || || ||| ||
Db        401 ACAGCCGTGAGCCGCCGCGCCTG 423


RESULT 9
CK000205
LOCUS       CK000205                788 bp    mRNA    linear   EST 26-NOV-2003
DEFINITION  AGENCOURT_16363613 NIH_MGC_220 Homo sapiens cDNA clone
            IMAGE:30707474 5', mRNA sequence.
ACCESSION   CK000205
VERSION     CK000205.1  GI:38526239
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Martin, University of Iowa
             cDNA Library Preparation: M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM1072  row: c  column: 03
            High quality sequence stop: 675.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30707474"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_220"
                     /note="Organ: mixed; Vector: pYX-Asc; Site_1: EcoRI;
                     Site_2: NotI; Library is oligo-dT primed and directionally
                     cloned Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according tomRNA size fraction, ligated with
                     EcoR I adaptor, digested with Not I and then cloned
                     directionally into pYX-Asc vector. Average insert size
                     0.5-1Kb. Adaptors 5'(AATTCGGCACGAGG)3' and 5'd
                     (CCTCGTGCCG)3'. 3' Linker sequence - GCGGCCGCTGAGAGCC T18.
                     Sequencing primers 3'end: T3 promoter primer 5'd
                     (ATTAACCCTCACTAAAGGGA)3'. 5' End: T7 promoter primer 5'd
                     (TAATACGACTCACTATAGGG)3'. Library was constructed in the
                     laboratory of M. Bento Soares. Average insert size 3-4kb
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             38.2%;  Score 139.4;  DB 7;  Length 788;
  Best Local Similarity   69.0%;  Pred. No. 1.2e-18;
  Matches  223;  Conservative    0;  Mismatches   91;  Indels    9;  Gaps    2;

Qy         52 GTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCT 111
              | || |||| || ||||| |||||||  | |   ||  ||||| ||||  | |||| |||
Db        106 GGGTTTTTTTTTTGTTTTGTTTTGTTTGTTTTGTTTTGTTTTTTTGAGACGGTGTCTCGCT 165

Qy        112 CCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTG 171
               | ||| | ||  |||||||||| || | ||||    ||| ||| |||||     |||||
Db        166 GCATCACCCAGGCTGGAGTGCAGTGGCGCAATCTCGCCTTCCTGCAGTCTCCGCCTCCTG 225

Qy        172 AGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACC 231
              |||||||| ||| || |||||||||| ||||| ||||||| ||||  | | ||    |||
Db        226 AGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTGCAGGCGCGCACACC 285

Qy        232 ATGCACAGCTGATTTTTAAATTTTTTTTGTAGA--------GATGGAGTTGCCCAGGCTG 283
              | || ||||| ||||||  ||||||  |  |||         ||   |||| ||||||||
Db        286 ACGCCCAGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCATCATGTTGACCAGGCTG 345

Qy        284 GTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGA-CCTCCCAAGTATCTTAGACT 342
              ||||  |||||||| | | |||||||||| || || || | ||||||||    ||  || |
Db        346 GTCTCAAACTCCTGACTTCAGGTGATCCACCCGCCTTCAGCCTCCCAACGTGCTGGGATT 405

Qy        343 ACAGATGCACTCCACCACGCTTG 365
              ||||  |    || || ||| ||
Db        406 ACAGCCGTGAGCCGCCGCGCCTG 428
```

```
RESULT 10
CD656882
LOCUS       CD656882                837 bp    mRNA    linear   EST 18-JUN-2003
DEFINITION  AGENCOURT_14538022 NIA Human H1 Embryonic Stem Cell cDNA Library
            (Long) Homo sapiens cDNA clone IMAGE:30422172 5', mRNA sequence.
ACCESSION   CD656882
VERSION     CD656882.1  GI:31898550
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 837)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Irene Ginis and Mahendra Rao, NIA
             cDNA Library Preparation: Yulan Piao and Minoru Ko
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC c            lone distribution information
            can be found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM501  row: c  column: 13
            High quality sequence stop: 660.
FEATURES             Location/Qualifiers
     source          1. .837
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30422172"
                     /tissue_type="Embryonic Stem cells"
                     /cell_line="WA01"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIA Human H1 Embryonic Stem Cell cDNA Library
                     (Long)"
                     /note="Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI;
                     This is a long-transcript enriched cDNA library (Genome
                     Res. 11: 1553-1558 (2001). [PMID: 11544199]) from WA01
                     cell line . Undifferentiated human ES cell line WA01/H1
                     was obtained from WiCell Research Institute, Inc.,
                     Madison, Wi, cultured according to their instructions, on
                     MEF feeders. They formed round colonies with defined edges
                     and were positive for alkaline phosphatase, SSEA-4, OCT3,
                     OCT4, REX1, UTF, TERT, SOX2, CX43 and CX45.  They are
                     negative for GATA2, GATA4, PDX1, NCAM, MSX1, FLT3, SSEA-1,
                     TUBB3, NES, GFAP, and EOMES. When confluent (18-10 days
                     after plating), the ES cells from 4 X 6cm dishes were
                     treated with 1 mg/ml collagenase, type IV
                     (Invitrogen/GIBCO) for 5-10 min and gently scraped off
                     with 5 ml pipette.  RNA was purified with TRIzol Reagent
                     from Invitrogen. Protocol ref:  Genome Res. 11: 1553-1558
                     (2001). [PMID:11544199]) Double-stranded cDNAs were
                     synthesized with an Oligo(dT) primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT-3'] from
                     3.4g of total RNA, treated with T4 DNA polymerase, and
                     purified by ethanol-precipitation. The cDNAs were ligated
                     to Lone-linker LL-Sal4, purified by phenol/chloroform
                     extraction, and separated from free linkers by
                     Centricon-100 column.  Then, the cDNAs were amplified by
                     long-range high fidelity PCR using Ex Taq polymerase
                     (Takara) with a primer Sal4-S for 25 cycles.  The products
                     were purified by phenol/chloroform extraction and
                     Centricon-100 column. The cDNAs were digested with SalI
                     and NotI enzymes and cloned into SalI/NotI site of
                     pCMV-SPORT6 plasmid vector. The average insert size is
                     about 3.6kb."
```

ORIGIN

Query Match 38.2%; Score 139.4; DB 6; Length 837;
Best Local Similarity 69.0%; Pred. No. 1.2e-18;
Matches 223; Conservative 0; Mismatches 91; Indels 9; Gaps 2;

```
Qy         52 GTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCT 111
              | || |||| || ||||| |||||||  | |   ||  ||||| ||||  | |||| |||
Db        215 GGGTTTTTTTTTTGTTTTGTTTTGTTGTTTTGTTTTGTTTTTTTGAGACGGTGTCTCGCT 274

Qy        112 CCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTG 171
               | ||| | ||  |||||||||| || | ||||    ||| ||| |||||     |||||
Db        275 GCATCACCCAGGCTGGAGTGCAGTGGCGCAATCTCGCCTTCCTGCAGTCTCCGCCTCCTG 334

Qy        172 AGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGCTGCTACC 231
              |||||||| ||| || |||||||||| ||||| ||||||| ||||  | | ||   |||
Db        335 AGTTCAAGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTGCAGGCGCGCACACC 394

Qy        232 ATGCACAGCTGATTTTTAAATTTTTTTTGTAGA--------GATGGAGTTGCCCAGGCTG 283
              | || ||||| ||||||  ||||||  |  |||          ||  |||| ||||||||
Db        395 ACGCCCAGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCATCATGTTGACCAGGCTG 454

Qy        284 GTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGA-CCTCCCAAGTATCTTAGACT 342
              ||||  |||||||| | | |||||||||| || || || | ||||||||     ||  || |
Db        455 GTCTCAAACTCCTGACTTCAGGTGATCCACCCGCCTTCAGCCTCCCAACGTGCTGGGATT 514

Qy        343 ACAGATGCACTCCACCACGCTTG 365
              ||||  |    || || ||| ||
Db        515 ACAGCCGTGAGCCGCCGCGCCTG 537
```

RESULT 11
H73550
LOCUS H73550 422 bp mRNA linear EST 31-OCT-1995
DEFINITION ys10h07.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:214429 5' similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION H73550
VERSION H73550.1 GI:1046609
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 422)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 487
High quality sequence stops: 368
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 487 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 368.
FEATURES Location/Qualifiers
source 1. .422
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:214429"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen 1NFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 38.1%; Score 139.2; DB 7; Length 422;
Best Local Similarity 69.5%; Pred. No. 1.5e-18;
Matches 221; Conservative 0; Mismatches 88; Indels 9; Gaps 2;

```
Qy         57 TTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCCGTC 116
              | | |||  |||  ||| | |  ||||  ||| |||||  |||| || |||||||| |||
Db         63 TGTCATTAATTTTTTTTTTCTTTTCTTTTTTTCTTTTTTGAGACAGAGTCTTGCTCTGTC 122

Qy        117 ATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAGTTC 176
              |   ||  |||||||||| ||||  ||| | ||| |||  || ||      ||||||||||
Db        123 ACTCAGGCTGGAGTGCAGTGGTGCGATCTTGGCTCACTACAGCCTCTGCCTCCTGAGTTC 182

Qy        177 AAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGC-TGCTACCATGC 235
              |||  || ||  || |||||| |||   ||||||| ||||| | ||||  || |||| ||
Db        183 AAGCTATTCTCATGTCTCAGCCTCCTGAGTAGCTGGGACTACAGGTGCACGCCACCACGC 242

Qy        236 ACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG--------AGTTGCCCAGGCTGGTCT 287
              |||  ||||||  ||||||  |  |||   ||          |||| ||||||||||||
Db        243 CTGGCTAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACTATGTTGGCCAGGCTGGTCT 302

Qy        288 TGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACTACAGA 347
              ||||||||  ||| | |||||||  ||||| |||  ||||||||     ||  || |||||
Db        303 CAAACTCCTGACCTCAAGTGATCCACCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGG 362

Qy        348 TGCACTCCACCACGCTTG 365
              ||    |||||||||  ||
Db        363 TGTGAGCCACCACGCCTG 380
```

RESULT 12
AW884394/c
LOCUS AW884394 384 bp mRNA linear EST 23-MAY-2000
DEFINITION QV3-OT0065-290300-137-a06 OT0065 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW884394
VERSION AW884394.1 GI:8046406
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 384)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research

```
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV3-OT0065-290
            300-137-a06&t3=2000-03-29&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 384.
FEATURES              Location/Qualifiers
     source           1. .384
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /dev_stage="Adult"
                      /clone_lib="OT0065"
                      /note="Organ: ovary; Vector: puc18; Site_1: SmaI; Site_2:
                      SmaI; A mini-library was made by cloning products derived
                      from  ORESTES PCR (U.S. Letters Patent application No.
                      196,716 -  Ludwig Institute for Cancer Research) profiles
                      into  the pUC 18 vector. Reverse transcription of tissue
                      mRNA and cDNA amplification were performed under low
                      stringency conditions."
ORIGIN

  Query Match             38.0%;  Score 138.8;  DB 2;  Length 384;
  Best Local Similarity   72.6%;  Pred. No. 1.9e-18;
  Matches  212;  Conservative    0;  Mismatches   67;  Indels   13;  Gaps    2;

Qy          87 TTATTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCAT 146
               ||  ||||||||| |||||||| |||| ||      ||    | |||||||  |||| ||||||
Db         331 TTCGTTTTAGAGACAGGGTCTCGCTCTGTTGCTCAGGCCGAAGTGCAGTGGTGCAATCAT 272

Qy         147 AGCTTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGT 206
               |||| |||| || ||  || ||||| | ||||| |||||| ||||||||||  ||||  ||
Db         271 AGCTCACTGCAGCCTCCAACTCCTGGGCTCAAGCGATCCTCCTGCCTCAGCCTCCCCAGT 212

Qy         207 AGCTGAGACTATATGTGCTGC----TACCATGCACAGCTGATTTTTAAATTTTTTTTGTA 262
               ||||||||||| ||  | |||      ||||||   ||| ||||||||||||||| |  |
Db         211 AGCTGAGACTACATAGGGTGCATGACCCCATGCCTGGCTAATTTTTAAATTTTTTGTAAA 152

Qy         263 GAGATGG---------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTC 313
               || | ||          |||||||||||| ||||  ||||| |||||| | |||||||||
Db         151 GACAGGGGTCTCACTATGTTGCCCAGGCTTGTCTCAAACTCTTGGCCTCATGTGATCCTC 92

Qy         314 CTGCGTTGACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGCTTG 365
                ||| |||  || ||||||| ||   |||| || ||  | ||||||| || |
Db          91 TTGCCTTGGTCTGCCAAGTAGCTGGTACTATAGGTGTGCACCACCACACTAG 40


RESULT 13
HSM806754/c
LOCUS       HSM806754               2198 bp    mRNA    linear   HTC 03-AUG-2004
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686A24102 (from clone DKFZp686A24102).
ACCESSION   BX641053
VERSION     BX641053.1  GI:34364972
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2198)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp686A24102) is available at
            the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in
            Berlin, Germany. Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686A24102
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES              Location/Qualifiers
     source           1. .2198
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="DKFZp686A24102"
                      /tissue_type="fetal kidney"
                      /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                      DH10B; sites SfiIA + SfiIB"
                      /dev_stage="fetal"
                      /note="unspliced mRNA"
ORIGIN

  Query Match             38.0%;  Score 138.8;  DB 3;  Length 2198;
  Best Local Similarity   69.0%;  Pred. No. 1.4e-18;
  Matches  225;  Conservative    0;  Mismatches   87;  Indels   14;  Gaps    2;

Qy          51 TGTGTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGC 110
               ||  | | |   | || |  | | |||  | ||| |||||||||| ||||  ||||||| |
Db        1246 TGACTGTCTCTATTTTATTTTATATTTTATTTTATTTTATTTTTTGAGACGGGGTCTTAC 1187

Qy         111 TCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCT 170
               || |||| | ||  |||||||||| ||||  ||| ||||| |||| || ||  || ||||
Db        1186 TCTGTCACCCAGGCTGGAGTGCAGTGGTGGGATCGTAGCTCACTGCAGCCTCAAACTCCT 1127

Qy         171 GAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTGC-TGCTA 229
               | | |||||  ||||| |||||||||| ||||| ||||||| ||||||| || | ||  |
Db        1126 GGGCTCAAGTAATCCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTATAGGTACATGTCA 1067

Qy         230 CCATGCACAGCTGATTTTTAAATTTTTTTTGTAG------------AGATGGAGTTGCC 276
               |||||| ||||| |||   ||| ||||||| | |                |  |  ||| |
Db        1066 CCATGCCCAGCTAATTAAAAAAATTTTTTTTTTGTATAGATCGGGTCTCACTGTATTGAC 1007

Qy         277 CAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCT 336
               |||||||||||| |||||||| || | | |||||||||| ||| ||||| ||    |
Db        1006 CAGGCTGGTCTTAAACTCCTGGTCTCAAGAGATCCTCCTGCCTTGGCCTCCTAATATGCC 947

Qy         337 TAGACTACAGATGCACTCCACCACGC 362
                  | |||||       |||||||||
Db         946 GGCATTACAGGCATGAGCCACCACGC 921


RESULT 14
CR749384
LOCUS       CR749384                8056 bp    mRNA    linear   HTC 19-AUG-2004
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686F02110 (from clone DKFZp686F02110).
ACCESSION   CR749384
VERSION     CR749384.1  GI:51476491
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 8056)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the
            Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA
```

sequencing consortium of the German Genome Project. This clone (DKFZp686F02110) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany. Please contact RZPD for ordering: http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686F02110 Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.

FEATURES Location/Qualifiers
source 1. .8056
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686F02110"
/tissue_type="fetal kidney"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host DH10B; sites SfiIA + SfiIB"
/dev_stage="fetal"
/note="genomic"

ORIGIN

Query Match 38.0%; Score 138.6; DB 3; Length 8056;
Best Local Similarity 69.0%; Pred. No. 1.2e-18;
Matches 223; Conservative 0; Mismatches 89; Indels 11; Gaps 2;

```
Qy         54 GTATTTTATTCTTTTGATTTGTTTAGTCTTACTTTATTTTTAGAGAAAGGGTCTTGCTCC 113
              || |||| || | ||  |||||||  | ||  ||| ||||| || |  | |||||||||
Db       4743 GTGTTTTTTTTTGTTTGTTTGTTTTTTGTTTTTTTTTTTTTTTGAAATGGAGTCTTGCTCT 4802

Qy        114 GTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGCTTACTGTAGTCTTGAATTCCTGAG 173
              |||||| ||| |||||| ||| ||     |||   ||| ||||||  ||      ||||| |
Db       4803 GTCATCCAGACTGGAGTACAGTGGCACTATCTCGGCTCACTGTAACCTCCGCCTCCTGGG 4862

Qy        174 TTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGCTGAGACTATATGTG-CTGCTACCA 232
              ||||||  || || ||||| |||| ||||| |||| || || || | |||  ||| ||||
Db       4863 TTCAAGCAATTCTCCTGCCCCAGCCTCCCAAGTAGTTGTGATTACAGGTGTGTGCCACCA 4922

Qy        233 TGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATGG----------AGTTGCCCAGGCT 282
              ||| | ||| ||||||  |||| ||| ||||||||||          |||| |||||||
Db       4923 TGCCCGGCTAATTTTTGTATTTCTTTAGTAGAGATGGGGTTTCACCACGTTGGCCAGGCT 4982

Qy        283 GGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTTGACCTCCCAAGTATCTTAGACT 342
              |||||  |||||||| ||| ||||||||| ||  | |   ||||||||     ||  || |
Db       4983 GGTCTCAAACTCCTGACCTCAGGTGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATT 5042

Qy        343 ACAGATGCACTCCACCACGCTTG 365
              | |||      || |||| | ||
Db       5043 ATAGAGTGGAACCGCCACACCTG 5065
```

RESULT 15
AI332671
LOCUS AI332671 360 bp mRNA linear EST 13-FEB-1999
DEFINITION qq33f08.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1934343 3' similar to contains Alu repetitive element;, mRNA sequence.
ACCESSION AI332671
VERSION AI332671.1 GI:4069230
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 360)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 831 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 253.

FEATURES Location/Qualifiers
source 1. .360
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1934343"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/clone_lib="Soares_NhHMPu_S1"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbHH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

ORIGIN

Query Match 37.9%; Score 138.4; DB 1; Length 360;
Best Local Similarity 73.4%; Pred. No. 2.3e-18;
Matches 207; Conservative 0; Mismatches 66; Indels 9; Gaps 2;

```
Qy         90 TTTTTAGAGAAAGGGTCTTGCTCCGTCATCTAGATTGGAGTGCAGCGGTGTAATCATAGC 149
              ||||| |||| ||||||| ||| || ||| |   |||||||||| ||  | |||| |||
Db          2 TTTTTGGAGACAGGGTCTCACTCTGTTATCCAAGCTGGAGTGCAGTGGCATGATCACAGC 61

Qy        150 TTACTGTAGTCTTGAATTCCTGAGTTCAAGAGATCCTTCTGCCTCAGCTTCCCAGGTAGC 209
              | |||| || ||||| |||||| ||||||| |||||| |  ||||||  |||   |||||
Db         62 TCACTGCAGCCTTGACTTCCTGGGTTCAAGTGATCCTCCCACCTCAGTCTCCTGAGTAGC 121

Qy        210 TGAGACTATATG-TGCTGCTACCATGCACAGCTGATTTTTAAATTTTTTTTGTAGAGATG 268
              || ||||| | |    | | ||||||| ||||| || ||||||||||| |  ||| | |
Db        122 TGGGACTACAGGCATGTACCACCATGCTCAGCTAATCTTTAAATTTTTTGTAGAGACAGG 181

Qy        269 G--------AGTTGCCCAGGCTGGTCTTGAACTCCTGGCCTGAGGTGATCCTCCTGCGTT 320
              |        ||||||||||||||||||||||||||    | | ||||| |||||| ||
Db        182 GTCTCACTTTGTTGCCCAGGCTGGTCTTGAACTCCTGAGTTCAAGTGATTCTCCTGCCTT 241

Qy        321 GACCTCCCAAGTATCTTAGACTACAGATGCACTCCACCACGC 362
              |||||||||    ||  || ||||| ||    ||||||| |
Db        242 GACCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCACAC 283
```

Search completed: August 30, 2005, 05:15:23
Job time : 2202 secs